



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 134603

TO: James Schultz
Location: REM/2D18/2C18
Art Unit: 1635
Friday, October 08, 2004

Case Serial Number: 09/721414

From: Mary Jane Ruhl
Location: Biotech-Chem Library
Remsen 1-A-62
Phone: 571-272-2524

maryjane.ruhl@uspto.gov

Search Notes

Examiner Schultz,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
Remsen 1-A-62
Ext. 22524

This Page Blank (uspto)

STIC-Biotech/ChemLib*134603*

From: Schultz, James
Sent: Wednesday, October 06, 2004 4:26 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search 09/721,414

Hello,
Could you please run standard nucleotide sequence searches on SEQ ID NOS: 9 and 16, which are 150 and 75 nucleotides long respectively. There are two sequences here because I cannot restrict between them--they are related because they have some region of homology.

Thanks,
Doug Schultz

James Douglas Schultz, PhD
AU 1635 (Biotechnology)
Patent Examiner
United States Patent and Trademark Office
(Office) REM 2D18
(Mail) REM 2C18
(571) 272-0763

RECEIVED
OCT - 6 2004
STIC/BIOTECH DIVISION
(STIC)

STAFF USE ONLY

Searcher: _____
Searcher Phone: 2- _____
Date Searcher Picked up: _____
Date Completed: _____
Searcher Prep/Rev. Time: _____
Online Time: _____

Type of Search

NA Sequence: # _____
AA Sequence: # _____
Structure: # _____
Bibliographic: _____
Litigation: _____
Patent Family: _____
Other: _____

Vendors and cost where applicable

STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other(Specify): _____

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2004, 01:43:05 ; Search time 769.667 Seconds

(without alignments)
4223.552 Million cell updates/sec

Title: US-09-721-414c-16

Perfect score: 75

Sequence: 1 gguguaacccaaggggna.....gaucucguaccgcagcca 75

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenBank1:*

1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pin:*
35: em_hlg_rtd:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_hlg_vrt:*
39: em_hlg_hum:*
40: em_hlg_mus:*
41: em_hlg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.2	86.9	75	12	SYNMBSC
2	64.4	85.9	75	1	ECOTRQ2AS
3	64.4	85.9	75	6	AX827247
4	64.4	85.9	183	6	AR389490
5	64.4	85.9	360	6	AR389250
6	64.4	85.9	644	6	AX497054
7	63.8	85.1	75	1	ECTGLN2
8	63.6	84.8	75	12	SYNMBSA
9	63.6	84.8	75	12	SYNMBSB
10	62.8	83.7	75	1	ECOTRQ2
11	62.8	83.7	75	1	ECOTRQ2A
12	62.8	83.7	75	1	ECTGLN2
13	62.8	83.7	75	1	ECTGLNB
14	62.8	83.7	75	6	176358
15	62.8	83.7	186	6	AR388871
16	62.8	83.7	1100	1	ECOTGCP
17	62.8	83.7	10137	1	AE015093
18	62.8	83.7	10627	1	AE000170
19	62.8	83.7	11871	1	AE005245
20	62.8	83.7	18276	1	D90706
21	62.8	83.7	22671	1	AE008727
22	62.8	83.7	240050	1	AL627267
23	62.8	83.7	262278	1	AP002552
24	62.8	83.7	290029	1	AE016980
25	62.8	83.7	300099	1	AE016841
26	62.8	83.7	300413	1	AE016757
27	60.4	80.5	75	12	SYNMBSD
28	59.6	79.5	11320	1	AE013723
29	59.6	79.5	243072	6	AX770904
30	59.6	79.5	258050	1	AJ414153
31	59.6	79.5	349318	1	BX571863
32	54.8	73.1	300272	1	AE017213
33	54.8	73.1	301838	1	AE017209
34	53.2	70.9	10192	1	AE006104
35	53.2	70.9	10596	1	AE015562
36	51.6	68.8	75	1	ECOTRQ1
37	51.6	68.8	75	6	176357
38	51.6	68.8	207	6	AR376242
39	51.6	68.8	228	6	AR389287
40	51.6	68.8	246	6	AR376457
41	51.6	68.8	357	6	AR389050
42	51.6	68.8	301191	1	AE017152
43	50	66.7	11933	1	U32836
44	50	66.7	110000	6	AR274513_16
45	48.4	64.5	711	6	BD092666

ALIGNMENTS

RESULT 1
LOCUS SYNMBSC
DEFINITION Synthetic amber suppressor Su(+)2-89 tRNA-Gln1 gene.
ACCESSION M13716
VERSION M13716.1 GI:207925
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

SYNMBSC 75 bp DNA linear SYN 26-APR-1995
Synthetic amber suppressor Su(+)2-89 tRNA-Gln1 gene.
M13716
M13716.1 GI:207925
amber suppressor; recombinant DNA; recombination; synthetic
sequence; termination suppressor; transfer RNA; transfer RNA
suppressor.
synthetic construct
artificial sequences.
1 (bases 1 to 75)
Yarus M., Cline S., Ratfery, L., Wier, P. and Bradley, D.
The translational efficiency of tRNA is a property of the anticodon

modified_base	54	/mod_base=p
ORIGIN	5'	end of mature trRNA.
Query Match	85.9%;	Score 64.4; DB 1; Length 75;
Best Local Similarity	73.0%;	Pred. No. 1.6e-12;
Matches	54;	Conservative 14; Mismatches 6; Indels 0; Gaps 0;
QY	2	GUGUUAUCCCAAGGGUAAAGGACCGGUAUUCUAAUACCGGCAUUCGAGUUGCAUCC 61
Db	2	GGGGTATTCGCAACGGGTAAAGGACCGGATTCTAATTCGGCAATTCGAGGTTCGAATCC 61
QY	62	UCGUACCGCAGCCA 75
Db	62	TCGTACCCCGACCA 75
RESULT 3		
AX827247		
LOCUS	AX827247	75 bp DNA linear PAT 12-DEC-2003
DEFINITION	Sequence 4 from patent WO03074719.	
ACCESSION	AX827247	
VERSION	AX827247.1	GI:39837336
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS	1	Rieping,M. and Siebel,C.N.
TITLE		Amino acid-producing bacteria and a process for preparing
JOURNAL		1-aminoacids
FEATURES		Patent: WO 03074719-A 4 12-SEP-2003;
source		Degussa AG (DE)
		Location/Qualifiers
	1..75	/organism="Escherichia coli"
		/mol_type="unassigned DNA"
	1..75	/db_xref="taxon:562"
		/note="subB-allele"
ORIGIN		
Query Match	85.9%;	Score 64.4; DB 6; Length 75;
Best Local Similarity	73.0%;	Pred. No. 1.6e-12;
Matches	54;	Conservative 14; Mismatches 6; Indels 0; Gaps 0;
QY	2	GUGUUAUCCCAAGGGUAAAGGACCGGUAUUCUAAUACCGGCAUUCGAGUUGCAUCC 61
Db	2	GGGGTATTCGCAACGGGTAAAGGACCGGATTCTAATTCGGCAATTCGAGGTTCGAATCC 61
QY	62	UCGUACCGCAGCCA 75
Db	62	TCGTACCCCGACCA 75
RESULT 4		
AR389490/c		
LOCUS	AR389490	183 bp DNA linear PAT 18-DEC-2003
DEFINITION	Sequence 6219 from patent US 6610836.	
ACCESSION	AR389490	
VERSION	AR389490.1	GI:40099441
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS	1	Unkown.
TITLE		Unclassified.
JOURNAL		1 (bases 1 to 183)
FEATURES		Bretton,G.L. and Osborne,M.
source		Nucleic acid amino acid sequences relating to Klebsiella pneumoniae
		for diagnostics and therapeutics
		Patent: US 6610836-A 6219 26-AUG-2003;
		Location/Qualifiers
	1..183	

SYNAMBSA	LOCUS	DEFINITION	ACCESSION
SYNAMBSA	75 bp	DNA linear	SYN 26-APR-1995
		suppressor Su(+2 tRNA-Gln1 gene.	
	M13714		

VERSION	M13714.1	GI:207923
KEYWORDS	amber suppressor; recombinant DNA; recombination; synthetic sequence; termination suppressor; transfer RNA; transfer RNA suppressor.	
SOURCE	synthetic construct	
ORGANISM	synthetic construct	
REFERENCE	1 (bases 1 to 75)	
AUTHORS	Yarus,M., Cline,S., Ratfery,L., Wier,P. and Bradley,D.	
TITLE	The translational efficiency of tRNA is a property of the anticodon arm	
JOURNAL	J. Biol. Chem. 261 (23), 10496-10505 (1986)	
MEDLINE	86278115	
PUBMED	3525546	
COMMENT	source text: Artificial gene DNA.	
FEATURES	Location/Qualifiers	
source	1..75	
	/organism="synthetic construct"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:32630"	
	1..75	
	/product="tRNA-Gln"	
	/note="codon recognized: UAG; synthetic Su(+)2 amber suppressor tRNA-Gln1"	
	/anticodon=(pos:33..35,aa:Gln)	
	26..27	
	/organism="Artificial gene"	
	39..40	
	/organism="Artificial gene"	
ORIGIN	-	
Query Match	84.8%; Score 63.6; DB 12; Length 75;	
Best Local Similarity	73.0%; Pred. No. 3.1e-12;	
Matches	54; Conservative 14; Mismatches 6; Indels 0; Gaps 0;	
OY	2 GUGUAUCCCAAGGCGUAAGGACCGAUAUCCGCGAUCGAGGUGCAUCC 61	
DB	2 GGGGATGCGCAAGCGGADAGGACCGAGTTCTAATTCGGCATTCGAGTTCGATTC 61	
OY	62 UCGUACCGCAGCCA 75	
	: :	
DB	62 TCGTACCCGAGCCA 75	
RESULT 9		
LOCUS	SYNAMSBS 75 bp DNA linear SYN 26-APR-1995	
SYNAMSBS		
DEFINITION	Synthetic amber suppressor Su(+)2-88 tRNA-Gln1 gene.	
ACCESSION	M13715	
VERSION	M13715.1	
KEYWORDS	MI3715.1 GI:207924	
	amber suppressor; recombinant DNA; recombination; synthetic sequence; termination suppressor; transfer RNA; transfer RNA suppressor.	
SOURCE	synthetic construct	
ORGANISM	synthetic construct	
REFERENCE	1 (bases 1 to 75)	
AUTHORS	Yarus,M., Cline,S., Ratfery,L., Wier,P. and Bradley,D.	
TITLE	The translational efficiency of tRNA is a property of the anticodon arm	
JOURNAL	J. Biol. Chem. 261 (23), 10496-10505 (1986)	
MEDLINE	86278115	
PUBMED	3525546	
COMMENT	source text: Artificial gene DNA.	
FEATURES	Location/Qualifiers	
source	1..75	
	/organism="synthetic construct"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:32630"	
	1..75	
	/product="tRNA-Gln"	
	/note="codon recognized: UAG; synthetic Su(+)2-88 amber suppressor tRNA-Gln1"	
	suppressor tRNA-Gln1"	
	suppressor tRNA-Gln1"	

misc_recomb	/antibiocond=(pos:33.35,aa:Gln)	
misc_recomb	26.27	
misc_recomb	/organism="Artificial gene"	
misc_recomb	39.40	
misc_recomb	/organism="Artificial gene"	
Query Match	84.8%; Score 63.6; DB 12; Length 75;	
Best Local Similarity	73.0%; Pred. No. 3,1e-12;	
Matches	54; Conservative 14; Mismatches 6; Indels 0; Gaps 0;	
Qy	62	UCCGACCGCAGCCA 75
Db	62	TTCGATCCCGACCA 75
RESULT 10		
ECOTR2		
LOCUS	ECOTR2	75 bp
DEFINITION	E. coli Gln-tRNA-2.	linear
ACCESSION	K00182	
VERSION	K00182.1	GI:174447
KEYWORDS	transfer RNA; transfer RNA suppressor; transfer RNA-Gln.	
SOURCE	Escherichia coli	
ORGANISM	Escherichia coli	
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.	
AUTHORS	1 (bases 1 to 75)	
TITLE	Yaniv,M. and Folik,W.R.	
JOURNAL	The nucleotide sequences of the two glutamine transfer ribonucleic acids from Escherichia coli	
MEDLINE	J. Biol. Chem. 250 (9), 3243-3253 (1975)	
PUBMED	75133562	
COMMENT	164464	
FEATURES	2 (bases 1 to 75)	
source	Bradley,D., Park,J.V. and Soll,L.	
source	tRNA 2Gln Su+2 mutants that increase amber suppression	
source	J. Bacteriol. 145 (2), 704-712 (1981).	
source	81117021	
source	7007345	
source	Original source text: E. coli K12 (strain W3110) tRNA [1] and Su+2 tRNA [2].	
source	Contributed on tape April 1983 by M.Sprinzl & D.H.Gaus; from their entry 0520 in Nucleic Acids Res. 11, R1-R54 (1983). [1] also determined the Gln-tRNA-1 sequence.	
source	Location/Qualifiers	
source	1..75	
source	/organism="Escherichia coli"	
source	/mol_type="tRNA"	
source	/db_xref="taxon:562"	
source	1..75	
source	/product="tRNA-Gln"	
source	/note="codon recognized: CAG; Gln-tRNA-2 (NAR: 0520)"	
source	8	/mod_base=s4u
source	17	/mod_base=gm
source	19	/mod_base=d
source	31	/mod_base=um
source	35	/mod_base=um
source	/note="g in wild type; a in Su+2 mutant (cta amber suppressor anticodon)"	
source	36	/mod_base=m2a
source	37	/mod_base=p
source	38	/mod_base=p


```

modified_base      /mod_base=p
                    53 /mod_base=t
modified_base      54 /mod_base=p
ORIGIN             5' end of mature tRNA.

Query Match
Best Local Similarity 83.7%; Score 62.8; DB 1; Length 75;
Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Oy 2 GUGGUAUCCCAAGGAGGUAAGGACCGGAUUCUAUCCGCAUUCGAGUUGCAUCC 61
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 GGGGATATGCCAAGCGGTAAGGACCGGATTCGATTCGGCATTCCAGGTTCAATCC 61
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 62 UCGUACCGCAGCCA 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 TCGTACCCAGCCA 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 11
ECOTR02A          75 bp tRNA linear BCT 15-APR-1994
LOCUS             E.coli Gln-tRNA-2.
DEFINITION        M10707
ACCESSION         M10707
VERSION           M10707.1 GI:174448
KEYWORDS          transfer RNA-Gln.
SOURCE            Escherichia coli
ORGANISM          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
                  Enterobacteriaceae; Escherichia.
REFERENCE         1 (bases 1 to 75)
AUTHORS           Inokuchi,H., Kodaira,M., Yamao,F. and Ozeki,H.
TITLE             Identification of transfer RNA suppressors in Escherichia coli. II.
                  Duplicate genes for tRNA2Gln
JOURNAL           J. Mol. Biol. 132 (4), 663-677 (1979)
MEDLINE           80117837
PUBMED            160950
COMMENT           Original source text: Escherichia coli tRNA.
FEATURES          location/Qualifiers
                    source
                    1..75
                    /organism="Escherichia coli"
                    /mol_type="tRNA"
                    /db_xref="taxon:562"
                    1..75
                    /product="tRNA-Gln"
                    /note="codon recognized: CAG"
                    modified_base
                    17 /mod_base=gm
                    modified_base
                    19 /mod_base=gm
                    modified_base
                    31 /mod_base=d
                    modified_base
                    31 /mod_base=um
                    modified_base
                    36 /mod_base=m2a
                    modified_base
                    37 /mod_base=p
                    modified_base
                    38 /mod_base=p
                    modified_base
                    54 /mod_base=p
                    modified_base
                    54 /mod_base=p
ORIGIN            5' end of mature tRNA.

Query Match
Best Local Similarity 83.7%; Score 62.8; DB 1; Length 75;
Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Oy 2 GUGGUAUCCCAAGGAGGUAAGGACCGGAUUCUAUCCGCAUUCGAGUUGCAUCC 61
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 GGGGATATGCCAAGCGGTAAGGACCGGATTCGATTCGGCATTCCAGGTTCAATCC 61
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 62 UCGUACCGCAGCCA 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 62 TCGTACCCAGCCA 75

RESULT 12
ECOTLN2
LOCUS             E.coli tRNA-Gln-2, and tRNA-Gln-2 UGA suppressor mutants su+2-am,
DEFINITION        su+2-oc and su+2-UGA.
ACCESSION         X55374 X55375 X55376 X55377
VERSION           X55374.1 GI:43057
KEYWORDS          suppressor tRNA; transfer RNA-Gln.
SOURCE            Escherichia coli
ORGANISM          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
                  Enterobacteriaceae; Escherichia.
REFERENCE         1 (bases 1 to 75)
AUTHORS           Inokuchi,H., Kondo,K., Yoshimura,M. and Ozeki,H.
TITLE             Mutant of the glutamine transfer RNA gene as UGA suppressor in
                  Escherichia coli
JOURNAL           Mol. Gen. Genet. 223 (3), 433-437 (1990)
MEDLINE           91101608
PUBMED            2270083
COMMENT           See X55374 for wild type. X55375, X55376 and X55377 for UGA
                  suppressor mutants.
FEATURES          location/Qualifiers
                    source
                    1..75
                    /organism="Escherichia coli"
                    /mol_type="genomic DNA"
                    /strain="K12"
                    /db_xref="taxon:562"
                    1..75
                    /product="tRNA-Gln"
                    /note="codon recognized: CAG; tRNA-2-Gln"
                    /anticodon="(pos:33..35,aa:Gln)"
                    gene
                    33 /gene="su+2-oc"
                    variation
                    33 /gene="su+2-oc"
                    /phenotype="UGA suppressor"
                    /replace="t"
                    gene
                    34 /gene="su+2-UGA"
                    variation
                    34 /gene="su+2-UGA"
                    /phenotype="UGA suppressor"
                    /replace="c"
                    gene
                    35 /gene="su+2-am"
                    variation
                    35 /gene="su+2-am"
                    /gene="su+2-am"
                    /phenotype="UGA suppressor"
                    /replace="a"
ORIGIN            5' end of mature tRNA.

Query Match
Best Local Similarity 83.7%; Score 62.8; DB 1; Length 75;
Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Oy 2 GUGGUAUCCCAAGGAGGUAAGGACCGGAUUCUAUCCGCAUUCGAGUUGCAUCC 61
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 GGGGATATGCCAAGCGGTAAGGACCGGATTCGATTCGGCATTCCAGGTTCAATCC 61
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 62 UCGUACCGCAGCCA 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 TCGTACCCAGCCA 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13
ECOTLN2
LOCUS             E.coli transfer RNA-Gln II.
DEFINITION        X68630
ACCESSION         X68630.1 GI:43059
VERSION           X68630.1 GI:43059

```

Db	2	GGGGTATCGCCAGCGGTAAGGACCGGATTCGTGATTCGGCATTCGAGGTTGCAATCC	61
QY	62	UCGUACCGCAGCCA	75
Db	62	TCGTACCCCGACCA	75
RESULT 15			
LOCUS	AR388871/c	186 bp	DNA
DEFINITION	Sequence 5600 from patent US 6610836.		Linear
ACCESSION	AR388871		
VERSION	AR388871.1	GI:40098605	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 186)		
AUTHORS	Bretton,G.L. and Osborne,M.		
TITLE	Nucleic acid amino acid sequences relating to <i>Klebsiella pneumoniae</i>		
JOURNAL	for diagnostics and therapeutics		
FEATURES	Patent: US 6610836-A 5600 26-AUG-2003;		
SOURCE	Location/Qualifiers		
	1..186		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Query Match	83.7%;	Score 62.8;	DB 6;
Best Local Similarity	71.6%;	Pred. No. 6,76-12;	Length 186;
Matches	53;	Conservative 14;	Mismatches 7;
			Indels 0;
			Gaps 0;
QY	2	GUGGUAUCCCCAAGGGGUAAGGACCGGAAUUCUAAUCCGGCAUUCGAGGUGCGAAUCC	61
Db	167	GGGGTATCGCCAGCGGTAAGGACCGGATTCGTGATTCGGCATTCGAGGTTGCAATCC	108
QY	62	UCGUACCGCAGCCA	75
Db	107	TCGTACCCCGACCA	94
Search completed: October 8, 2004, 04:44:50			
Job time : 773.667 secs			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 7, 2004, 22:35:09 ; Search time 191.667 Seconds
(without alignments)
1662.338 Million cell updates/sec

Title: US-09-721-414C-16

Perfect score: 75

Sequence: 1 ggugguauccccaaggggga.....gauccuguaacgcagcca 75

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : N_Geneseq_29Jan04:*

1: geneseqn19808:*
2: geneseqn19908:*
3: geneseqn20008:*
4: geneseqn20018:*
5: geneseqn20018:*
6: geneseqn20028:*
7: geneseqn20038:*
8: geneseqn20038:*
9: geneseqn20038:*
10: geneseqn20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	100.0	75	4 AAC85943	AAC85943 cRNA. 8/
2	71.8	95.7	75	4 AAC85935	AAC85935 V2 varian
3	64.4	85.9	644	6 AAL47112	AAL47112 Peptide P
4	64	85.3	76	4 AAC85940	AAC85940 cRNA DNA
5	62.8	83.7	2007	5 AAS94249	AAS94249 DNA encod
6	62.8	83.7	2383	5 AAS79656	AAS79656 DNA encod
7	62.8	83.7	3169	5 AAS87641	AAS87641 DNA encod
8	62.8	83.7	3175	5 AAS92254	AAS92254 DNA encod
9	60.8	81.1	76	9 ADC16642	ADC16642 Aminoacyl
10	59.8	79.7	100	7 ACD68842	ACD68842 E. coli K
11	59.6	79.5	110000	7 ACD68842	Continuation (20 o
12	59.6	79.5	243072	7 ACD68842	Continuation (20 o
13	53	70.7	160	4 AAC85937	AAC85937 Photorhab
14	51.6	68.8	100	4 ACD68857	AAC85937 H2 leu ca
15	51.6	68.8	100	4 ACD68857	ACD68857 E. coli K
16	51.6	68.8	3999	5 AAS82444	AAS82444 DNA encod
17	51	68.0	73	4 AAC85936	AAC85936 V3 varian
18	50	66.7	165	7 ACF69090	ACF69090 Photorhab
19	50	66.7	110000	2 AAT42063_16	Continuation (17 o
20	48.4	64.5	711	2 AAX14366	AAX14366 H. pylori
21	47.8	63.7	73	4 AAC85934	AAC85934 VI varian
22	46.2	61.6	100	7 ACD68849	ACD68849 E. coli K
23	44.8	59.7	156	4 AAC85938	AAC85938 DI-leu ca

24	43.8	58.4	74	9 ADC16643	ADC16643 Aminoacyl
25	42.6	56.8	110000	8 ADB12064_06	Continuation (7 of
26	42.2	56.3	110000	8 ADB12064_04	Continuation (5 of
27	41.6	55.5	308	2 AAV78694	AAV78694 Staphyloc
28	40.8	54.4	110000	3 AAF22303_2	Continuation (3 of
29	40.6	54.1	987	2 AAX13475	AAX13475 Enterococ
30	40.6	54.1	987	6 AAS99270	ABS99270 Enterococ
31	40.4	53.9	399	6 AEN91909	ABN91909 Staphyloc
32	40.4	53.9	400	2 AAV78012	AAV78012 Staphyloc
33	40.4	53.9	1051	2 AAV77885	AAV77885 Staphyloc
34	40.4	53.9	3736	4 AAH55027	AAH55027 S. epider
35	40.4	53.9	110000	6 ABO69245_22	Continuation (23 o
36	40.4	53.9	110000	6 ABO69245_22	Continuation (23 o
37	40.4	53.9	213251	6 ABO67193	ABG67193 Listeria
38	39	52.0	300	2 AAO89663	AAQ89663 ochre sup
39	39	52.0	716	2 AAV78101	AAV78101 Staphyloc
40	39	52.0	2006	2 AAO89689	AAQ89689 pfg1. 3/2
41	38.8	51.7	1398	6 AEN70725	ABN70725 Streptoc
42	38.8	51.7	110000	6 AEN71527_11	Continuation (12 o
43	37.8	50.4	377	2 AAV78326	AAV78326 Staphyloc
44	37.8	50.4	400	2 AAV78104	AAV78104 Staphyloc
45	37.4	49.9	72	6 ABE77131	ABK77131 Bacillus

ALIGNMENTS

RESULT 1	AAC85943	AAC85943 standard; RNA; 75 BP.
ID	AAC85943	
AC	AAC85943;	
XX		
DT	22-AUG-2001	(first entry)
XX		
DE	oRNA.	
XX		
KM	Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;	
XX	amino acid; aminoacylation; ss.	
OS	Synthetic.	
XX		
FT	Key	Location/Qualifiers
FT	stem_loop	10..24
FT	stem_loop	/*tag= a
FT	stem_loop	/label= D-loop
FT	stem_loop	26..42
FT	stem_loop	/*tag= b
FT	stem_loop	/label= AC-loop
FT	D_loop	43..47
FT	D_loop	/*tag= c
FT	D_loop	/label= V-loop
FT	stem_loop	48..64
FT	stem_loop	/*tag= d
FT	stem_loop	/label= T-loop
XX		
XX	WO200138582-A1.	
XX	31-MAY-2001.	
XX	22-NOV-2000; 2000WO-US032184.	
XX	24-NOV-1999; 99US-0167331P.	
XX	28-JUN-2000; 2000US-0214382P.	
XX	(UNYV) UNIV NEW YORK STATE RES FOUND.	
XX	Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzaghi N;	
XX	WPI; 2001-381295/40.	
XX	Catalytic RNA having aminoacylating activity comprises a tRNA-like and a	
XX	ribozyme domain and is useful to produce non-natural amino acids for	
PT		

PT biomedical and therapeutic purposes.
XX
PS Disclosure; Fig 3; 38pp; English.
XX
CC This sequence represents orthogonal tRNA (tRNA), a catalytic RNA
CC molecule, designated a cis-aminoacylating RNA, which comprises a tRNA-
CC like and a ribozyme domain, and which has the ability to selectively
CC aminoacylate its own 3' terminus with specific amino acids. The ribozyme
CC domain has the catalytic activity and also confers amino acid specificity
CC for aminoacylation. The catalytic RNA molecules are used to aminoacylate
CC tRNA-like molecules in cis or in trans for the production of proteins
CC containing non-natural amino acids that may be of use for biomedical or
CC therapeutic purposes. The catalytic RNA molecules were constructed by
CC applying selection to a randomly synthesised RNA pool. By attaching the
CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
CC contacting the complex with a substrate molecule (natural or non-natural
CC amino acid), self-aminoacylating RNA molecules were identified
XX
SQ Sequence 75 BP; 17 A; 21 C; 22 G; 0 T; 15 U; 0 Other;
Query Match 100.0%; Score 75; DB 4; Length 75;
Best Local Similarity 100.0%; Pred. No. 2,3e-20;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GUGGUAUCCCCCAAGGAGGACCGGUAUUCUAAUCCGCAUUCGAGUUCGAUUC 60
Db 1 GUGGUAUCCCCCAAGGAGGAGGACCGGUAUUCUAAUCCGCAUUCGAGUUCGAUUC 60
QY 61 CUCGUACCGGAGCCA 75
Db 61 CUCGUACCGGAGCCA 75
RESULT 2
AAC85935
ID AAC85935 standard; RNA; 75 BP.
XX
AC AAC85935;
XX
DT 22-AUG-2001 (first entry)
XX
DE V2 variant of ocRNA.
XX
KM Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
KM amino acid; aminoacylation; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT stem_loop 10..24 /*tag= a
FT /label= D-loop
FT 26..42 /*tag= b
FT /label= AC-loop
FT D_loop 43..47 /*tag= c
FT /label= V-loop
FT stem_loop 49..64 /*tag= d
FT /label= T-loop
XX
PN WO200138562-A1.
XX
PD 31-MAY-2001.
XX
PE 22-NOV-2000; 2000WO-US032184.
XX
PR 24-NOV-1999; 99US-0167331P.
XX 26-JUN-2000; 2000US-0214382P.
XX (UJNY) UNIV NEW YORK STATE RES FOUND.
XX

PI Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzaghi N;
XX WPI; 2001-381295/40.
XX
DR
XX
PT Catalytic RNA having aminoacylating activity comprises a tRNA-like and a
PT ribozyme domain and is useful to produce non-natural amino acids for
PT biomedical and therapeutic purposes.
XX
PS Claim 3; Page 36; 38pp; English.
XX
CC The sequences given in AAC85923-36 are catalytic RNA molecules,
CC designated cis-aminoacylating RNA, which comprise a tRNA-like and a
CC ribozyme domain, and which have the ability to selectively amino-acylate
CC their own 3' terminus with specific amino acids. The ribozyme domain has
CC the catalytic activity and also confers amino acid specificity for
CC aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
CC -like molecules in cis or in trans for the production of proteins
CC containing non-natural amino acids that may be of use for biomedical or
CC therapeutic purposes. The catalytic RNA molecules were constructed by
CC applying selection to a randomly synthesised RNA pool. By attaching the
CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
CC contacting the complex with a substrate molecule (natural or non-natural
CC amino acid), self-aminoacylating RNA molecules were identified
XX
SQ Sequence 75 BP; 17 A; 22 C; 21 G; 0 T; 15 U; 0 Other;
Query Match 95.7%; Score 71.8; DB 4; Length 75;
Best Local Similarity 97.3%; Pred. No. 4.8e-19;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GUGGUAUCCCCCAAGGAGGAGGACCGGUAUUCUAAUCCGCAUUCGAGUUCGAUUC 60
Db 1 GUGGUAUCCCCCAAGGAGGAGGACCGGUAUUCUAAUCCGCAUUCGAGUUCGAUUC 60
QY 61 CUCGUACCGGAGCCA 75
Db 61 CUCGUACCGGAGCCA 75
RESULT 3
AAL47112
ID AAL47112 standard; DNA; 644 BP.
XX
AC AAL47112;
XX
DT 20-AUG-2002 (first entry)
XX
DE Peptide presentation method related vector fragment.
XX
KM Peptide presentation; host cell surface; antibody isolation;
KM epitope mapping; bacteria; vector; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT primer_bind 1..20 /*tag= a
FT /bound_moiety= "Primer SupE2-Ecoup"
FT /note= "binds primer shown in AAL47113"
FT primer_bind 626..644 /*tag= b
FT /bound_moiety= "Primer SupE2-Mlu-1o"
FT /note= "binds primer shown in AAL47116"
XX
PN WO200234906-A2.
XX
PD 02-MAY-2002.
XX
PE 25-OCT-2001; 2001WO-DE004009.
XX
PR 26-OCT-2000; 2000DE-01053224.
XX (UYGE-) UNIV GEORG AUGUST GOETTINGEN.
XX

```

XX  Kolmar H, Christmann A, Wentzel A;
XX
XX  WPI: 2002-435621/46.
XX
XX  Cell-surface presentation of peptides or proteins, useful e.g. for
XX  isolating monospecific antibodies, comprises expression of a fusion
XX  sequence with truncated intimin, in bacteria.
XX
XX  Example 1; Fig 6A; 44pp; German.
XX
CC  The present invention relates to a method for presenting peptides or
CC  proteins on the surface of host bacteria. The method can be used to
CC  isolate, from polyclonal mixtures, monospecific antibodies that bind
CC  selectively to a particular surface-bound polypeptide, or more generally
CC  any specifically binding partners, and for epitope mapping. The present
CC  sequence is a vector fragment described in the exemplification of the
XX  invention
XX
SQ  Sequence 644 BP; 148 A; 172 C; 157 G; 167 T; 0 U; 0 Other;
SQ
Query Match      85.9%; Score 64.4; DB 6; Length 644;
Best Local Similarity 73.0%; Pred. No. 8.6e-16;
Matches 54; Conservative 14; Mismatches 6; Indels 0; Gaps 0;
QY  2 GUGGUAUCCCAAGGAGGUAAGGAGACCGGAAUUCGUAUCCGAGUUCGAGUUCGAAUCC 61
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db  346 GGGGATGCGCCAAAGCGGTAAGGACACCGGATTTCTAATTCGCGATTCCAGTTCGAATCC 405
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY  62 UCGUACCGCAGACCA 75
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db  406 TCGTAACCCAGACCA 419
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

RESULT 4	
ID	AAC85940/c
XX	AAC85940 standard; DNA; 76 BP.
XX	
AC	AAC85940;
XX	
DT	22-AUG-2001 (first entry)
XX	
DE	ocrRNA DNA template for generation of catalytic RNA's.
XX	
KW	Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
KM	amino acid; aminoacylation; ss.
XX	
OS	Synthetic.
XX	
PN	WC200138582-A1.
XX	
PD	31-MAY-2001.
XX	
PF	22-NOV-2000; 2000WO-US032184.
XX	
PR	24-NOV-1999; 99US-016731P.
PR	28-JUN-2000; 2000US-0214382P.
XX	
PA	(UNYNY) UNIV NEW YORK STATE RES FOUND.
XX	
PI	Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzaghi N;
XX	
DR	WPI; 2001-381295/40.
XX	
PT	Catalytic RNA, having aminoacylating activity comprises a tRNA-like and a
PT	ribozyme domain and is useful to produce non-natural amino acids for
XX	biomedical and therapeutic purposes.
XX	
PS	Example 1; Page 31; 38pp; English.
XX	
CC	The sequences given in AAC85939-42 were used in the construction of a
CC	pool of RNA molecules for screening of aminoacylation activity.
CC	Randomised sequences were generated and attached to a tRNA-like molecule.

[illegible]

XX	RESULT 5
XX	AAS94249
ID	AAS94249 standard; cDNA; 2007 BP.
XX	
AC	AAS94249;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #30053.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PE	30-MAR-2001; 2001WO-US008631.
XX	
PR	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
FI	Drmnac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB; ABG30062.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.
XX	
PS	Claim 1; SEQ ID NO 30053; 103bp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II).
CC	sequences. (I) is useful as hybridisation probes, polymerase chain
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC	and in recombinant production of (II). The polynucleotides are also used
CC	in diagnostics as expressed sequence tags for identifying expressed
CC	genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

Query Match 83.7%; Score 62.8; DB 5; Length 2007;
 Best Local Similarity 71.6%; Pred. No. 5.2e-15;
 Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GUGGUAUCCCAAGGGGUAAGGACCGGAUUCUAAUCCGCGAUUCCAGGUCCGAUCC 61
 Db 885 GGGGTATCGCCAGCGGTAAAGGACCGGATTCTGATTCGCGATTCCGAGTTGGAATCC 944
 Qy 62 UCGUACCGCAGCCA 75
 Db 945 TCGTACCCAGCCA 958

RESULT 6
 ID AAS79656 standard; cDNA; 2383 BP.
 AC AAS79656;
 DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #15460.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US008631.
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG15469.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 PS Claim 1; SEQ ID NO 15460; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

Query Match 83.7%; Score 62.8; DB 5; Length 2383;
 Best Local Similarity 71.6%; Pred. No. 5.5e-15;
 Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GUGGUAUCCCAAGGGGUAAGGACCGGAUUCUAAUCCGCGAUUCCAGGUCCGAUCC 61
 Db 885 GGGGTATCGCCAGCGGTAAAGGACCGGATTCTGATTCGCGATTCCGAGTTGGAATCC 944
 Qy 62 UCGUACCGCAGCCA 75
 Db 945 TCGTACCCAGCCA 958

RESULT 7
 ID AAS87641 standard; cDNA; 3169 BP.
 AC AAS87641;
 DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #23445.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US008631.
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG23454.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 PS Claim 1; SEQ ID NO 23445; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SO Sequence 3169 BP, 812 A, 841 C, 779 G, 737 T, 0 U, 0 Other;

Query Match 83.7%; Score 62.8; DB 5; Length 3169;

Best Local Similarity 71.6%; Pred. No. 5.9e-15;

Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

QY 2 GUGGUAUCCCGCAAGGGGUAAGGACCGCAUUCUAAUCCGCAUUCGAGUCCGAUCC 61

DB 2047 GGGGATATGCCAAGGCGTAAGGACCGGATTCGATTCGCGGATTCGAGGTTGCAATCC 2106

QY 62 UCGUACCGGACCA 75

DB 2107 TCGTACCCGACCA 2120

RESULT 8

AAS92254

ID AAS92254 standard; cDNA; 3175 BP.

AC AAS92254;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #28058.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US008631.

PF 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR P-PSDB; ABG28067.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

PS Claim 1; SEQ ID NO 28058; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SO Sequence 3175 BP, 811 A, 849 C, 779 G, 736 T, 0 U, 0 Other;

Query Match 83.7%; Score 62.8; DB 5; Length 3175;

Best Local Similarity 71.6%; Pred. No. 5.9e-15;

Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

QY 2 GUGGUAUCCCGCAAGGGGUAAGGACCGCAUUCUAAUCCGCAUUCGAGUCCGAUCC 61

DB 2053 GGGGATATGCCAAGGCGTAAGGACCGGATTCGATTCGCGGATTCGAGGTTGCAATCC 2112

QY 62 UCGUACCGGACCA 75

DB 2113 TCGTACCCGACCA 2126

RESULT 9

ADCI6642

ID ADCI6642 standard; tRNA; 76 BP.

AC ADCI6642;

DT 18-DEC-2003 (first entry)

DE Aminoacylation RNA molecule related cRNA.

XX ribozyme; aminoacylate; tRNA; non-cognate; catalytic RNA molecule; cis;

KW aminoacylation; trans; proteomic; ss.

XX Unidentified.

OS WO2003070740-A1.

PN 28-AUG-2003.

PD 18-FEB-2003; 2003WO-US005007.

PF 15-FEB-2002; 2002US-0357424P.

XX (UNYV) UNIV NEW YORK STATE RES FOUND.

PA Suga H, Murakami H, Saito H;

PI WPI; 2003-748198/70.

XX New polynucleotide, useful for preparing peptides containing non-cognate

PT amino acids, encodes ribozyme that can aminoacylate tRNA with such amino

PT acids.

PS Disclosure; SEQ ID NO 2; 85pp; English.

XX The invention relates to a novel polynucleotide comprising a sequence

CC encoding a ribozyme that can aminoacylate tRNA with a non-cognate amino
CC acid. Ribozymes encoded by the polynucleotide of the invention are used
CC to prepare polypeptides that contain non-cognate, including non-natural,
CC amino acids. The invention more specifically provides catalytic RNA
CC molecules having cis aminoacylation activity with a catalytic and
CC aminoacylation domain, or an RNA molecule with trans aminoacylation
CC activity with only a catalytic domain. The products of the invention are
CC potentially useful for biomedical and therapeutic use, e.g. for probing
CC the structure and function of proteins; preparation of peptide libraries
CC and in proteomics. This polynucleotide sequence represents an cRNA
CC structure relating to the RNA molecule with aminoacylation activity of
CC the invention.

XX SQ Sequence 76 BP; 16 A; 23 C; 23 G; 0 T; 14 U; 0 Other;

Query Match 81.1%; Score 60.8; DB 9; Length 76;

Best Local Similarity 96.1%; Pred. No. 1,4e-14;

Matches 73; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GGUGUAVUCCCAAGGAGGAGGAGCGGATUCU-AAUCCGCGAUCGAGUUCGAAU 59

Db 1 GGUGUAVUCCCAAGGAGGAGGAGCGGATUCUCCGAAUCCGCGAUCGAGUUCGAAU 60

QY 60 CCUCGUACCGGAGCCA 75

Db 61 CCUCGUACCGGAGCCA 76

RESULT 10

ACD68842/c

XX ID ACD68842 standard; DNA; 100 BP.

ACD68842;

XX 18-SEP-2003 (first entry)

DE E. coli K12 MG1655 biochip probe SEQ ID 112.

XX Blochip; gene expression; gut; diagnostic; detection; probe; ss.

XX Becherichia coli.

XX EPI260592-A1.

XX 27-NOV-2002.

XX 17-MAY-2001; 2001EP-00112179.

XX 17-MAY-2001; 2001EP-00112179.

XX (MWGB-) MWG-BIOTECH AG.

XX Donner H, Drescher B, Huber A, Weber J;

XX WPI; 2003-241155/24.

XX Claim 3; Page 28; 2004PP; German.

CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single

CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population densities for the various strains. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention

XX SQ Sequence 100 BP; 23 A; 27 C; 29 G; 21 T; 0 U; 0 Other;

Query Match 79.7%; Score 59.8; DB 7; Length 100;

Best Local Similarity 70.4%; Pred. No. 3.7e-14;

Matches 50; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

QY 2 GGUGUAVUCCCAAGGAGGAGGAGCGGATUCUAAUCCGCGAUCGAGUUCGAAUCC 61

Db 71 GGUGUAVUCCCAAGGAGGAGGAGCGGATUCUAAUCCGCGAUCGAGUUCGAAUCC 12

QY 62 UCGUACCGGAG 72

Db 11 TCGTACCCGAG 1

RESULT 11

ACF67367 19

Continuation (20 of 57) of ACF67367 from base 1900001 (Photophobus luminescens nucleotic

WP Sequence Split into 57 fragments LOCUS ACF67367 Accession ACF67367

WP Fragment Name Begin End

WP ACF67367_00 1 110000

WP ACF67367_01 10001 21000

WP ACF67367_02 20001 31000

WP ACF67367_03 30001 41000

WP ACF67367_04 40001 51000

WP ACF67367_05 50001 61000

WP ACF67367_06 60001 71000

WP ACF67367_07 70001 81000

WP ACF67367_08 80001 91000

WP ACF67367_09 90001 101000

WP ACF67367_10 100001 111000

WP ACF67367_11 110001 121000

WP ACF67367_12 120001 131000

WP ACF67367_13 130001 141000

WP ACF67367_14 140001 151000

WP ACF67367_15 150001 161000

WP ACF67367_16 160001 171000

WP ACF67367_17 170001 181000

WP ACF67367_18 180001 191000

WP ACF67367_19 190001 201000

WP ACF67367_20 200001 211000

WP ACF67367_21 210001 221000

WP ACF67367_22 220001 231000

WP ACF67367_23 230001 241000

WP ACF67367_24 240001 251000

WP ACF67367_25 250001 261000

WP ACF67367_26 260001 271000

WP ACF67367_27 270001 281000

WP ACF67367_28 280001 291000

WP ACF67367_29 290001 301000

WP ACF67367_30 300001 311000

WP ACF67367_31 310001 321000

WP ACF67367_32 320001 331000

WP ACF67367_33 330001 341000

WP ACF67367_34 340001 351000

WP ACF67367_35 350001 361000

WP ACF67367_36 360001 371000

WP ACF67367_37 370001 381000

WP ACF67367_38 380001 391000

WP ACF67367_39 390001 401000

WP ACF67367_40 400001 411000

WP ACF67367_41 410001 421000


```
WP ACF67367_42 420001 4310000
WP ACF67367_43 430001 4410000
WP ACF67367_44 440001 4510000
WP ACF67367_45 450001 4610000
WP ACF67367_46 460001 4710000
WP ACF67367_47 470001 4810000
WP ACF67367_48 480001 4910000
WP ACF67367_49 490001 5010000
WP ACF67367_50 500001 5110000
WP ACF67367_51 510001 5210000
WP ACF67367_52 520001 5310000
WP ACF67367_53 530001 5410000
WP ACF67367_54 540001 5510000
WP ACF67367_55 550001 5610000
WP ACF67367_56 560001 5648894

Query Match
Best Local Similarity 68.9%; Score 59.6; DB 7; Length 110000;
Matches 51; Conservative 14; Mismatches 9; Indels 0; Gaps 0;

QY 2 GUGUAUCCCCCAAGGCGUAAGGACCGGAUUCUAAUCCGCAUUCGAGUUCGAUCC 61
Db 86440 GGGGATATGCCCAAGCGGTAAAGCACCGGATTCGATTCGCGATTCGCCAGTTCCGATCC 86499
QY 62 UCGUACCGCAGCCA 75
Db 86500 TGGTACCCCGACCA 86513
```

```
RESULT 12
ACF65382
ID ACF65382 standard; DNA; 243072 BP.
```

```
AC ACF65382;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens nucleotide sequence #35.
XX
KM Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KM detection; food; gene expression; plant; animal; microorganism; toxin;
KM antibiotic; biopesticide; virulence factor; disease model; plague;
KM whooping cough; gene; de.
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-1B003040.
XX
PR 07-FEB-2001; 2001PR-00001659.
XX
PA (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Tsourit S, Glaeser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 1; SEQ ID NO 35; 1205bp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
```

```
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens genes
XX
SQ Sequence 243072 BP; 70273 A; 57885 C; 49427 G; 65487 T; 0 U; 0 Other;

Query Match
Best Local Similarity 79.5%; Score 59.6; DB 7; Length 243072;
Matches 51; Conservative 14; Mismatches 9; Indels 0; Gaps 0;

QY 2 GUGUAUCCCCCAAGGCGUAAGGACCGGAUUCUAAUCCGCAUUCGAGUUCGAUCC 61
Db 39182 GGGGATATGCCCAAGCGGTAAAGCACCGGATTCGATTCGCGATTCGCCAGTTCCGATCC 39241
QY 62 UCGUACCGCAGCCA 75
Db 39242 TGGTACCCCGACCA 39255
```

```
RESULT 13
```

```
AAC85937
ID AAC85937 standard; RNA; 160 BP.
```

```
AC AAC85937;
XX
DT 22-AUG-2001 (first entry)
XX
DB H2 leu catalytic RNA.
XX
KM Catalytic RNA; cis-aminocylating RNA; tRNA-like domain; ribozyme domain;
KM amino acid; aminocylation; ss.
XX
OS Synthetic.
XX
PN WO200138582-A1.
XX
PD 31-MAY-2001.
XX
PF 22-NOV-2000; 2000WO-US032184.
XX
PR 24-NOV-1999; 99US-0167331P.
XX 28-JUN-2000; 2000US-0214382P.
XX
PA (UNIV ) UNIV NEW YORK STATE RES FOUND.
XX
PI Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzagui N;
XX
DR WPI; 2001-381295/40.
XX
PT Catalytic RNA having aminocylating activity comprises a tRNA-like and a
XX PT ribozyme domain and is useful to produce non-natural amino acids for
XX PT biomedical and therapeutic purposes.
XX
PS Example 5; Page 35; 38pp; English.
XX
CC The sequences given in AAC85937-38 are catalytic RNA molecules,
CC designated cis-aminocylating RNA, which comprise a tRNA-like and a
CC ribozyme domain, and which have the ability to selectively amino-acylate
CC their own 3' termini with specific amino acids. These catalytic RNA's
CC are leucine specific cis aminocylating RNA's. The ribozyme domain has
CC the catalytic activity and also confers amino acid specificity for
```

CC aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
 CC -like molecules in cis or in trans for the production of proteins
 CC containing non-natural amino acids that may be of use for biomedical or
 CC therapeutic purposes. The catalytic RNA molecules were constructed by
 CC applying selection to a randomly synthesised RNA pool. By attaching the
 CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
 CC contacting the complex with a substrate molecule (natural or non-natural
 CC amino acid), self- aminoacylating RNA molecules were identified
 XX

SO Sequence 160 BP, 34 A; 37 C; 49 G; 0 T; 40 U; 0 Other;

Query Match Best Local Similarity 70.7%; Score 53; DB 4; Length 160;

Matches 64; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 11 CCAAGGUGAAGGACCGGATUCUAACCGGCAUCCGAGUUCGAUCCUGGACCGC 70
 DB 97 CCAAGGUGAAGGACCGGATUCUAACCGGCAUCCGAGUUCGAUCCUGGACCGC 155

OY 71 AGCCA 75
 DB 156 AGCCA 160

RESULT 14
 ACD68857/c
 ID ACD68857 standard; DNA; 100 BP.

AC ACD68857;

DT 18-SEP-2003 (first entry)

DE E. coli K12 MG1655 biochip probe SEQ ID 127.

KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.

OS Escherichia coli.

PN EP1260592-A1.

PD 27-NOV-2002.

PF 17-MAY-2001; 2001EP-00112179.

PR 17-MAY-2001; 2001EP-00112179.

PA (MWGB-) MWG-BIOTECH AG.

PI Donner H, Drescher B, Huber A, Weber J;

DR WPI; 2003-241155/24.

PT Biochip containing probes complementary with open reading frames in
 PT Escherichia coli K12, useful for detecting gene expression and expression
 PT patterns.

PS Claim 3; Page 30; 2004p; German.

XX This invention describes a novel biochip comprising probe spots, each
 CC containing many identical probes. The probes are nucleotide sequences of
 CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
 CC least one includes a segment of at least 20 bases identical with, or
 CC complementary to, a segment of an open reading frame (orf) of Escherichia
 CC coli K12. The biochip is used for specific detection of gene expression
 CC in K12 and for determining the gene expression pattern, e.g. for
 CC diagnostic determination of which E. coli strains are present in the gut,
 CC and to determine the effects of e.g. growth media on gene expression. The
 CC biochip provides a comprehensive analysis of possible detection of the K12
 CC genome, with simultaneous analysis of many different genes with a single
 CC device, and comparison of gene expression between K12 and its mutants or
 CC other E. coli strains in a single experiment. Apart from qualitative and
 CC quantitative information about gene expression, it also allows
 CC measurements of population densities for the various strains. The use of

CC synthetic oligonucleotides for preparation of probes allows free
 CC variation in probe length and ensures high purity (and thus selectivity,
 CC reactivity and reproducibility); also synthetic probes are generally
 CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
 CC AC81540 represent oligonucleotide probes used with the biochip described
 CC in the invention
 XX

SO Sequence 100 BP, 23 A; 27 C; 28 G; 22 T; 0 U; 0 Other;

Query Match Best Local Similarity 68.8%; Score 51.6; DB 7; Length 100;

Matches 48; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

OY 2 GUGGUAUCCCAAGGUGAAGGACCGGATUCUAACCGGCAUCCGAGUUCGAUCC 61
 DB 77 GGGGTAATCGCCAGGCGTAAGGACCGGTTTTCATACCGGCAATTCCTGTTGAAATCC 18
 OY 62 UCGUACCGCAGCCA 75
 DB 17 AGGTATCCCGACCA 4

RESULT 15
 ACD68856/c
 ID ACD68856 standard; DNA; 100 BP.

AC ACD68856;

DT 18-SEP-2003 (first entry)

DE E. coli K12 MG1655 biochip probe SEQ ID 126.

KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.

OS Escherichia coli.

PN EP1260592-A1.

PD 27-NOV-2002.

PF 17-MAY-2001; 2001EP-00112179.

PR 17-MAY-2001; 2001EP-00112179.

PA (MWGB-) MWG-BIOTECH AG.

PI Donner H, Drescher B, Huber A, Weber J;

DR WPI; 2003-241155/24.

PT Biochip containing probes complementary with open reading frames in
 PT Escherichia coli K12, useful for detecting gene expression and expression
 PT patterns.

PS Claim 3; Page 30; 2004p; German.

XX This invention describes a novel biochip comprising probe spots, each
 CC containing many identical probes. The probes are nucleotide sequences of
 CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
 CC least one includes a segment of at least 20 bases identical with, or
 CC complementary to, a segment of an open reading frame (orf) of Escherichia
 CC coli K12. The biochip is used for specific detection of gene expression
 CC in K12 and for determining the gene expression pattern, e.g. for
 CC diagnostic determination of which E. coli strains are present in the gut,
 CC and to determine the effects of e.g. growth media on gene expression. The
 CC biochip provides a comprehensive analysis of possible detection of the K12
 CC genome, with simultaneous analysis of many different genes with a single
 CC device, and comparison of gene expression between K12 and its mutants or
 CC other E. coli strains in a single experiment. Apart from qualitative and
 CC quantitative information about gene expression, it also allows
 CC measurements of population densities for the various strains. The use of
 CC synthetic oligonucleotides for preparation of probes allows free
 CC variation in probe length and ensures high purity (and thus selectivity,

CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACP68731 to
CC AC081540 represent oligonucleotide probes used with the biochip described
CC in the invention
CY

SD Sequence 100 BP; 23 A; 27 C; 28 G; 22 T; 0 U; 0 Other;

Query Match	68.8%	Score 51.6;	DB 7;	Length 100;
Best Local Similarity	64.9%;	Pred. No. 7.8e-11;		
Matches	48;	Conservative	12;	Mismatches 14;
			Indels	0;
			Gaps	0;

QY 2 GUGGUAUCCCCAAGGGGUAAAGGCACCCGAUUCUUAUACC GGCGCAUCCGAGGUTCCGAUCC 61
|||::||| |||::||| |||::||| |||::||| |||::||| |||::||| |||::||| |||::|||
Dd 77 GGGGTATCGCCCAAGCGGTAAAGCACCCGGTTTGTATACACGGCATTCCCTGTTGTAATCC 18

77 GGGGATCGCCAGCGGTAGGCAACGGTTTGTATACCGGCATTCCCTGGTTCGAATCC 18

QY 62 UCGUACCGCAGCCA 75

Db 17 AGGTACCCACGCCA 4

Search completed: October 8, 2004, 02:53:11
Job time : 194.667 secs

Large blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2004, 02:13:20 : Search time 38.6667 Seconds
(without alignments)
1076.413 Million cell updates/sec

Title: US-09-721-414C-16

Perfect score: 75

Sequence: 1 GUGGUAUCCCAAGGAGGUA.....GAUCCUGUACCGAGCA 75

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/PTUS_COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	64.4	85.9	183	4	US-09-489-039A-6219 Sequence 6219, Ap
C 2	64.4	85.9	360	4	US-09-489-039A-5979 Sequence 5979, Ap
C 3	62.8	83.7	75	1	US-08-242-098-28 Sequence 28, Appl
C 4	62.8	83.7	186	4	US-09-489-039A-5600 Sequence 5600, Ap
C 5	51.6	68.8	75	1	US-08-242-098-27 Sequence 27, Appl
C 6	51.6	68.8	207	4	US-09-543-681A-1248 Sequence 1248, Ap
C 7	51.6	68.8	228	4	US-09-489-039A-6016 Sequence 6016, Ap
C 8	51.6	68.8	246	4	US-09-543-681A-1463 Sequence 1463, Ap
C 9	51.6	68.8	357	4	US-09-489-039A-5779 Sequence 5779, Ap
C 10	50	66.7	1830121	4	US-09-557-884-1 Sequence 1, Appl
C 11	50	66.7	1830121	4	US-09-643-990A-1 Sequence 1018, Ap
C 12	46.8	62.4	288	4	US-09-543-681A-1018 Sequence 1421, Ap
C 13	46.8	62.4	288	4	US-09-543-681A-1421 Sequence 4383, Ap
C 14	41.6	55.5	308	1	US-08-242-098-25 Sequence 25, Appl
C 15	40.6	54.1	72	1	US-09-134-000C-892 Sequence 892, Appl
C 16	40.6	54.1	306	4	US-09-134-001C-1172 Sequence 1372, Ap
C 17	40.4	53.9	399	4	US-08-956-171E-3701 Sequence 3701, Ap
C 18	40.4	53.9	400	4	US-08-956-171E-3574 Sequence 3574, Ap
C 19	40.4	53.9	1051	4	US-08-242-098-32 Sequence 32, Appl
C 20	39	52.0	300	1	US-08-242-098-10 Sequence 10, Appl
C 21	39	52.0	716	4	US-08-956-171E-3790 Sequence 3790, Ap
C 22	39	52.0	2006	1	US-08-242-098-36 Sequence 36, Appl
C 23	39	52.0	2006	1	US-08-242-098-36 Sequence 36, Appl
C 24	38.8	51.7	77	1	US-08-242-098-15 Sequence 15, Appl
C 25	37.8	50.4	377	4	US-08-956-171E-4015 Sequence 4015, Ap
C 26	37.8	50.4	400	4	US-08-956-171E-3793 Sequence 3793, Ap
C 27	37.4	49.9	72	1	US-08-242-098-23 Sequence 23, Appl

28	37.4	49.9	72	1	US-08-242-098-31 Sequence 31, Appl
C 29	37.4	49.9	400	4	US-08-956-171E-4067 Sequence 4067, Ap
C 30	37.4	49.9	461	4	US-08-956-171E-3626 Sequence 3626, Ap
C 31	37.4	49.9	2456	4	US-08-961-527-213 Sequence 213, Appl
C 32	37.4	49.9	6591	4	US-08-956-171E-3114 Sequence 3114, Ap
C 33	37.4	49.9	11915	4	US-08-961-527-96 Sequence 96, Appl
C 34	37.2	49.6	75	1	US-08-242-098-29 Sequence 29, Appl
C 35	37.2	49.6	640681	4	US-09-790-098-1 Sequence 1, Appl
C 36	35.8	47.7	1230025	4	US-09-198-452A-1 Sequence 1, Appl
C 37	35.6	47.5	297	4	US-09-328-352-526 Sequence 526, Appl
C 38	35.6	47.5	297	4	US-09-328-352-526 Sequence 609, Appl
C 39	34.8	46.4	72	1	US-08-242-098-24 Sequence 24, Appl
C 40	34.6	46.1	718	4	US-08-956-171E-3762 Sequence 3762, Ap
C 41	34.4	45.9	72	1	US-08-242-098-12 Sequence 12, Appl
C 42	34.4	45.9	72	1	US-08-242-098-13 Sequence 13, Appl
C 43	34.2	45.6	72	1	US-08-242-098-14 Sequence 14, Appl
C 44	34.2	45.6	72	1	US-08-242-098-17 Sequence 17, Appl
C 45	34.2	45.6	72	1	US-08-242-098-18 Sequence 18, Appl

ALIGNMENTS

```

RESULT 1
US-09-489-039A-6219/c
; Sequence 6219, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 6219
; LENGTH: 183
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6219

Query Match      85.9%; Score 64.4; DB 4; Length 183;
Best Local Similarity 73.0%; Pred. No. 4.9e-17;
Matches 54; Conservative 14; Mismatches 6; Indels 0; Gaps 0;

QY      2 GUGGUAUCCCAAGGAGGUAAGGUAUCCGAGCAUCCGAGUUGCAUCC 61
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      77 GGGGTAATGCCAAGCGGTAAGGACCGATTTATTCGGCATTCGAGTTCAATCC 18
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      62 UGUAUCCGAGCA 75
      :||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      17 TCGTATCCGAGCA 4
      :||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 2
US-09-489-039A-5979/c
; Sequence 5979, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 5979
; LENGTH: 360
; TYPE: DNA

```

ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5979

Query Match 85.9%; Score 64.4; DB 4; Length 360;
Best Local Similarity 73.0%; Pred. No. 6e-17;
Matches 54; Conservative 14; Mismatches 6; Indels 0; Gaps 0;

QY 2 GUGUAUCCCAAGGAGGAGCGGUAUCCGGAUUCGAGGUGGUAUCC 61
DB 282 GGGGTATGCCAAGCGGTAAAGCAGCGATTCTATTCCGGCATTCGAGTTGCAATCC 223

QY 62 UCGUACCGCAGCCA 75
DB 222 TCGTACCCAGCCA 209

RESULT 3
US-08-242-098-28
Sequence 28, Application US/08242098
Patent No. 5691185

GENERAL INFORMATION:
APPLICANT: DICKELY, Françoise
APPLICANT: JOHANSEN, Eric

APPLICANT: NILSSON, Dan
APPLICANT: HANSEN, Egon

TITLE OF INVENTION: Lactic Acid Bacterial Suppressor Mutants
TITLE OF INVENTION: and Their Use as Selective Markers and as Means of
Containment in Lactic Acid Bacteria

NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA

ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,098
FILING DATE: 13-MAY-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,390
FILING DATE: 08-OCT-1993
ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30307/141/PLVI

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399

TELEX: 904136
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:

LENGTH: 75 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-242-098-28

Query Match 83.7%; Score 62.8; DB 1; Length 75;
Best Local Similarity 71.6%; Pred. No. 1.7e-16;
Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

QY 2 GUGUAUCCCAAGGAGGAGCGGUAUCCGGAUUCGAGGUGGUAUCC 61
DB 2 GGGGTATGCCAAGCGGTAAAGCAGCGATTCTATTCCGGCATTCGAGTTGCAATCC 61

QY 62 UCGUACCGCAGCCA 75
DB 62 TCGTACCCAGCCA 75

RESULT 4
US-09-489-039A-5600/c
Sequence 5600, Application US/09489039A
Patent No. 6610836

GENERAL INFORMATION:
APPLICANT: Gary Breton et. al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001

CURRENT FILING DATE: 2000-01-27
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 5600
LENGTH: 186
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5600

Query Match 83.7%; Score 62.8; DB 4; Length 186;
Best Local Similarity 71.6%; Pred. No. 2.2e-16;
Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

QY 2 GUGUAUCCCAAGGAGGAGCGGUAUCCGGAUUCGAGGUGGUAUCC 61
DB 167 GGGGTATGCCAAGCGGTAAAGCAGCGATTCTATTCCGGCATTCGAGTTGCAATCC 108

QY 62 UCGUACCGCAGCCA 75
DB 107 TCGTACCCAGCCA 94

RESULT 5
US-08-242-098-27
Sequence 27, Application US/08242098
Patent No. 5691185

GENERAL INFORMATION:
APPLICANT: DICKELY, Françoise
APPLICANT: JOHANSEN, Eric

APPLICANT: NILSSON, Dan
APPLICANT: HANSEN, Egon

TITLE OF INVENTION: Lactic Acid Bacterial Suppressor Mutants
TITLE OF INVENTION: and Their Use as Selective Markers and as Means of
Containment in Lactic Acid Bacteria

NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA

ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,098
FILING DATE: 13-MAY-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,390
FILING DATE: 08-OCT-1993
ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30307/141/PLVI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-242-098-27

Query Match
Best Local Similarity 68.8%; Score 51.6; DB 1; Length 75;
Matches 48; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 GUGUUAUCCCAAGGGGUAAGGACCGGAUUCUAAUCCGGCAUUCGAGUUCGAUCC 61
DB 2 GGGGTATCGCCAAAGCGGTAAAGCACCGGTTTGATACCGGCAATCCGTTGCAATCC 61

QY 62 UGUACCCGACGCCA 75
DB 62 AGTACCCGACGCCA 75

RESULT 6
US-09-543-681A-1248/c
Sequence 1248, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 1248
LENGTH: 207
TYPE: DNA
ORGANISM: Proteus mirabilis
US-09-543-681A-1248

Query Match
Best Local Similarity 68.8%; Score 51.6; DB 4; Length 207;
Matches 47; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

QY 2 GUGUUAUCCCAAGGGGUAAGGACCGGAUUCUAAUCCGGCAUUCGAGUUCGAUCC 61
DB 82 GGGGTATCGCCAAAGCGGTAAAGCACCGGTTTGATCCTGGCATTCGAGTTCAATCC 23

QY 62 UGUACCCGACGCCA 75
DB 22 TGTACCCGACGCCA 9

RESULT 7
US-09-489-039A-6016/c
Sequence 6016, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 6016
LENGTH: 228
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6016

Query Match
Best Local Similarity 68.8%; Score 51.6; DB 4; Length 228;
Matches 48; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 GUGUUAUCCCAAGGGGUAAGGACCGGAUUCUAAUCCGGCAUUCGAGUUCGAUCC 61
DB 205 GGGGTATCGCCAAAGCGGTAAAGCACCGGTTTGATACCGGCAATCCGTTGCAATCC 146

QY 62 UGUACCCGACGCCA 75
DB 145 AGTACCCGACGCCA 132

RESULT 8
US-09-543-681A-1463/c
Sequence 1463, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 1463
LENGTH: 246
TYPE: DNA
ORGANISM: Proteus mirabilis
US-09-543-681A-1463

Query Match
Best Local Similarity 68.8%; Score 51.6; DB 4; Length 246;
Matches 47; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

QY 2 GUGUUAUCCCAAGGGGUAAGGACCGGAUUCUAAUCCGGCAUUCGAGUUCGAUCC 61
DB 113 GGGGTATCGCCAAAGCGGTAAAGCACCGGTTTGATCCTGGCATTCGAGTTCAATCC 54

QY 62 UGUACCCGACGCCA 75
DB 53 TGTACCCGACGCCA 40

RESULT 9
US-09-489-039A-5779/c
Sequence 5779, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 5779
LENGTH: 357
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5779

Query Match 68.8%; Score 51.6; DB 4; Length 357;
Best Local Similarity 64.9%; Pred. No. 1e-11;
Matches 48; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 GUGUAUCCCAAGGAGGUAAGGACCGGAUUCUAAUCCGGAUUCGAGUUGCAUCC 61
DB 210 GGGGATATGCCAAGCGGTAAAGGACCGGTTTGTATACCGGCAATTCCTGGTGAATCC 151

QY 62 UCGUACCGCAGCCA 75
DB 150 AGTATACCCAGCCA 137

RESULT 10

US-09-557-884-1/c
Sequence 1, Application US/09557884
Patent No. 6506581

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884

FILING DATE: 25-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102

FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8439

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Query Match 66.7%; Score 50; DB 4; Length 1830121;
Best Local Similarity 62.2%; Pred. No. 5.8e-10;
Matches 46; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 2 GUGUAUCCCAAGGAGGUAAGGACCGGAUUCUAAUCCGGAUUCGAGUUGCAUCC 61
DB 1693458 GGGGATATGCCAAGCGGTAAAGGACCGGTTTGTATCTCAGCATTCCTAGGTGAATCC 1693399

QY 62 UCGUACCGCAGCCA 75
DB 1693398 TAGTATACCCAGCCA 1693385

RESULT 11
US-09-643-990A-1/c

Sequence 1, Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann

Mark D. Adams

Owen White

Hamilton O. Smlth

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Query Match 66.7%; Score 50; DB 4; Length 1830121;
Best Local Similarity 62.2%; Pred. No. 5.8e-10;
Matches 46; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 2 GUGUAUCCCAAGGAGGUAAGGACCGGAUUCUAAUCCGGAUUCGAGUUGCAUCC 61
DB 1693458 GGGGATATGCCAAGCGGTAAAGGACCGGTTTGTATCTCAGCATTCCTAGGTGAATCC 1693399

QY 62 UCGUACCGCAGCCA 75
DB 1693398 TAGTATACCCAGCCA 1693385

RESULT 12

US-09-543-681A-1018/c

Sequence 1018, Application US/09543681A

Patent No. 6605709

GENERAL INFORMATION:

APPLICANT: GARY BRETON

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

FILE REFERENCE: 2709.1002-001

CURRENT APPLICATION NUMBER: US/09/543,681A

CURRENT FILING DATE: 2000-04-05

PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1998-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 1018
LENGTH: 288
TYPE: DNA
ORGANISM: Proteus mirabilis
US-09-543-681A-1018

Query Match 62.4%; Score 46.8; DB 4; Length 288;
Best Local Similarity 62.2%; Pred. No. 8.6e-10;
Matches 46; Conservative 11; Mismatches 17; Indels 0; Gaps 0;

QY 2 GUGUUAUCCCAAGGAGGUAAGGACCGAUUUAUCCGGAUUCGAGUUGCAUCC 61
DB 110 GGGGATGCGCAAGGGTAAGGACCGAGTTGATCCTGGCATACCTGTTGAAATCC 51
QY 62 UCGUACCGCAGCCA 75
DB 50 AGGTACCCGACGCA 37

RESULT 13
US-09-543-681A-1421/c
Sequence 1421, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETTON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 1421
LENGTH: 288
TYPE: DNA
ORGANISM: Proteus mirabilis
US-09-543-681A-1421

Query Match 62.4%; Score 46.8; DB 4; Length 288;
Best Local Similarity 62.2%; Pred. No. 8.6e-10;
Matches 46; Conservative 11; Mismatches 17; Indels 0; Gaps 0;

QY 2 GUGUUAUCCCAAGGAGGUAAGGACCGAUUUAUCCGGAUUCGAGUUGCAUCC 61
DB 240 GGGGATGCGCAAGGGTAAGGACCGAGTTGATCCTGGCATACCTGTTGAAATCC 181
QY 62 UCGUACCGCAGCCA 75
DB 180 AGGTACCCGACGCA 167

RESULT 14
US-08-956-171E-4383
Sequence 4383, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunisch
GIL H. Choi
Patrick S. Dillon
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland

COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 4383:
SEQUENCE CHARACTERISTICS:
LENGTH: 308 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4383:
US-08-956-171E-4383

Query Match 55.5%; Score 41.6; DB 4; Length 308;
Best Local Similarity 56.8%; Pred. No. 1.2e-07;
Matches 42; Conservative 11; Mismatches 21; Indels 0; Gaps 0;

QY 2 GUGUUAUCCCAAGGAGGUAAGGACCGAUUUAUCCGGAUUCGAGUUGCAUCC 61
DB 211 GGGGATGCGCAAGGGTAAGGACCGAGTTGATCCTGGCATACCTGTTGAAATCC 270
QY 62 UCGUACCGCAGCCA 75
DB 271 ANGTAGCCGACGCA 284

RESULT 15
US-08-242-098-25
Sequence 25, Application US/08242098
Patent No. 5691185
GENERAL INFORMATION:
APPLICANT: DICKELLY, Francoise
APPLICANT: JOHANSEN, Eric
APPLICANT: NILSSON, Dan
APPLICANT: HANSEN, Egon
APPLICANT: STROMAN, Per
TITLE OF INVENTION: Lactic Acid Bacterial Suppressor Mutants
TITLE OF INVENTION: and their Use as Selective Markers and as Means of
CONTAINMENT in Lactic Acid Bacteria
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,098
FILING DATE: 13-MAY-1994

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/133,390
; FILING DATE: 08-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30307/141/PLVI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-242-098-25

```

```

Query Match      54.1%; Score 40.6; DB 1; Length 72;
Best Local Similarity 56.3%; Pred. No. 1.9e-07;
Matches 40; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY      2 GUGGUAUCCCCCAAGGAGGUAAGGACCGGAUUCUAAUCCGAGCAUUCGAGGUGCAUCC 61
      |||:|:|||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      2 GAGGTGTAGCCCAACGCGTAAAGCGACGACTTGACTCCGCGATTGTAAGTTGAAATCC 61
      :|||:|:|||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY      62 UCGUACCGCAG 72
      :|||:|:|||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      62 TACCACCTCAG 72
      :|||:|:|||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

Search completed: October 8, 2004, 04:47:01
 Job time : 44.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2004, 02:42:05 : Search time 197.333 Seconds
(without alignment)
1926.778 Million cell updates/sec

Title: US-09-721-414C-16

Perfect score: 75
Sequence: 1 gguguaucacccaaggsggga.....gaucucgacgagcca 75

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	64.4	85.9	75	US-10-347-484C-4	Sequence 4, Appli
2	64.4	85.9	644	US-10-415-165-16	Sequence 16, Appli
3	60.8	81.1	76	US-10-369-036B-2	Sequence 2, Appli
4	50	66.7	1830121	US-10-329-960-1	Sequence 1, Appli
5	50	66.7	1830121	US-10-329-960-1	Sequence 1, Appli
6	48.4	64.5	711	US-09-882-227-159	Sequence 159, App
7	43.8	58.4	74	US-10-369-036B-3	Sequence 3, Appli
8	41.6	55.5	308	US-08-781-986A-4383	Sequence 4383, Ap
9	41.6	55.5	308	US-10-329-624-4383	Sequence 4383, Ap
10	40.6	54.1	987	US-09-070-927A-538	Sequence 538, App
11	40.4	53.9	400	US-08-781-986A-3701	Sequence 3701, Ap
12	40.4	53.9	400	US-10-329-624-3701	Sequence 3701, Ap
13	40.4	53.9	1051	US-08-781-986A-3574	Sequence 3574, Ap
14	40.4	53.9	1051	US-10-329-624-3574	Sequence 3574, Ap

C 15	40.4	53.9	213251	16	US-10-398-221-6	Sequence 6, Appli
C 16	40.4	53.9	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
C 17	39	52.0	716	8	US-08-781-986A-3790	Sequence 3790, Ap
C 18	39	52.0	716	13	US-10-329-624-3790	Sequence 3790, Ap
C 19	38.2	50.9	2731748	17	US-10-297-465A-1	Sequence 1, Appli
C 20	37.8	50.4	377	8	US-08-781-986A-4015	Sequence 4015, Ap
C 21	37.8	50.4	377	13	US-10-329-624-4015	Sequence 4015, Ap
C 22	37.8	50.4	400	8	US-08-781-986A-3793	Sequence 3793, Ap
C 23	37.8	50.4	400	13	US-10-329-624-3793	Sequence 3793, Ap
C 24	37.4	49.9	72	9	US-09-974-300-4422	Sequence 4422, Ap
C 25	37.4	49.9	73	9	US-09-974-300-4432	Sequence 4432, Ap
C 26	37.4	49.9	75	9	US-09-974-300-4436	Sequence 4436, Ap
C 27	37.4	49.9	75	9	US-09-974-300-4431	Sequence 4431, Ap
C 28	37.4	49.9	400	8	US-08-781-986A-4067	Sequence 4067, Ap
C 29	37.4	49.9	400	13	US-10-329-624-4067	Sequence 4067, Ap
C 30	37.4	49.9	461	8	US-08-781-986A-3626	Sequence 3626, Ap
C 31	37.4	49.9	461	13	US-10-329-624-3626	Sequence 3626, Ap
C 32	37.4	49.9	2456	13	US-10-158-844-213	Sequence 213, App
C 33	37.4	49.9	6591	8	US-08-781-986A-3114	Sequence 3114, Ap
C 34	37.4	49.9	6591	13	US-10-329-624-3114	Sequence 3114, Ap
C 35	37.4	49.9	11915	13	US-10-158-844-96	Sequence 96, Appli
C 36	37.2	49.6	640681	9	US-09-790-988-1	Sequence 1, Appli
C 37	35.8	47.7	1230025	16	US-10-289-762-1	Sequence 1, Appli
C 38	34.6	46.1	718	8	US-08-781-986A-3762	Sequence 3762, Ap
C 39	34.6	46.1	718	13	US-10-329-624-3762	Sequence 3762, Ap
C 40	34.2	45.6	684707	16	US-10-398-221-9	Sequence 9, Appli
C 41	34.2	45.6	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
C 42	34	45.3	253	8	US-08-781-986A-4512	Sequence 4512, Ap
C 43	34	45.3	253	13	US-10-329-624-4512	Sequence 4512, Ap
C 44	32.6	43.5	531	17	US-10-437-963-3482	Sequence 3482, Ap
C 45	31.8	42.4	362	17	US-10-632-117-24	Sequence 24, Appli

ALIGNMENTS

```
RESULT 1
US-10-347-484C-4
; Sequence 4, Application US/10347484C
; Publication No. US20040082040A1
; GENERAL INFORMATION:
; APPLICANT: Degussa AG
; APPLICANT: Rieping, Mechthild
; APPLICANT: Siebel, Nicole
; TITLE OF INVENTION: Amino acid-producing bacteria and a process for preparing L-amino
; FILE REFERENCE: 7601/80186
; CURRENT APPLICATION NUMBER: US/10/347, 484C
; CURRENT FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 75
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: tRNA
; LOCATION: (1)..(75)
; OTHER INFORMATION: supe-A11ele
US-10-347-484C-4
Query Match 85.9%; Score 64.4; DB 17; Length 75;
Best Local Similarity 73.0%; Pred. No. 1.6e+14;
Matches 54; Conservative 14; Mismatches 6; Gaps 0;
```

RESULT 2

US-10-415-165-16
Sequence 16, Application US/10415165
Publication No. US20040106118A1
GENERAL INFORMATION:
APPLICANT: Georg-August-University, Göttingen
TITLE OF INVENTION: Method for Exposing Peptides and Polypeptides on the Cell Surface
TITLE OF INVENTION: of Bacteria
FILE REFERENCE: 031001744a
CURRENT APPLICATION NUMBER: US/10/415,165
CURRENT FILING DATE: 2003-10-15
PRIOR APPLICATION NUMBER: PPT/DE01/04009
PRIOR FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: DE 100 53 224.1
PRIOR FILING DATE: 2000-10-26
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.2
SEQ ID NO 16
LENGTH: 644
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: SupE-Eco
LOCATION: (1)..(644)
US-10-415-165-16

Query Match 85.9%; Score 64.4; DB 17; Length 644;
Best Local Similarity 73.0%; Pred. No. 2,2e-14;

Matches 54; Conservative 14; Mismatches 6; Indels 0; Gaps 0;

QY 2 GUGGUAUCCCAAGGAGGUAAGGACCGGUAUUCUAAUCCGCAUUCGAGGUGCAUUC 61
DB 346 GGGGTAATCGCCAAAGCGGTAAGGACCGGATTTCTAATTCGCGATTCGAGGTTGCAATCC 405

QY 62 UCGUACCGCAGCCA 75
DB 406 TCGTACCCCAAGCCA 419

RESULT 3

US-10-369-036B-2
Sequence 2, Application US/10369036B
Publication No. US20030228593A1
GENERAL INFORMATION:
APPLICANT: Suga, Hiroaki et al.
TITLE OF INVENTION: Ribozymes with broad tRNA aminoacylation activity
FILE REFERENCE: 11520.0230
CURRENT APPLICATION NUMBER: US/10/369,036B
CURRENT FILING DATE: 2003-02-18
PRIOR APPLICATION NUMBER: 60/357,424
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 61
SEQ ID NO 2
LENGTH: 76
TYPE: RNA
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: synthesized
US-10-369-036B-2

Query Match 81.1%; Score 60.8; DB 16; Length 76;
Best Local Similarity 96.1%; Pred. No. 3.6e-13;

Matches 73; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GGUGUAUCCCAAGGAGGUAAGGACCGGUAUUCU-AAAUCCGCAUUCGAGGUGUCCAAU 59
DB 1 GGUGUAUCCCAAGGAGGUAAGGACCGGUAUUCGCAUUCGAGGUGUCCAAU 60

QY 60 CCUCGUACCGAGCCA 75
DB 61 CCUCGUACCGAGCCA 76

RESULT 4

US-10-329-960-1/c
Sequence 1, Application US/10329960
Publication No. US20030099277A1
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fragr
TITLE OF INVENTION: Thereof, and Uses Thereof
FILE REFERENCE: P186P1
CURRENT APPLICATION NUMBER: US/10/329,960
CURRENT FILING DATE: 2003-01-02
PRIOR APPLICATION NUMBER: US 09/643,990
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: US 08/487,429
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/426,787
PRIOR FILING DATE: 1995-04-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (129298)..(129298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4416)..(4416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

```
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
```

```
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)
```

```
Query Match 66.7% Score 50; DB 15; Length 1830121;
Best Local Similarity 62.2% Pred. No. 2.1e-08;
Matches 46; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 2 GUGUAUCCCGAAGGGGUAAGGACCGAUUCAAUAUCCGGAUCCGAGUUGCAUCC 61
DB 1693458 GGGGTATCGCCAGCGGTAAAGCACTGGGTTTGATCTCAGCATCTTGAAGTTCGAATCC 1693399
QY 62 UGUACCGCAGGCA 75
DB 1693398 TAGTACCCAGCCA 1693385
```

```
RESULT 5
US-10-329-670-1/c
```

FEATURE:	NAME/key: misc.feature
LOCATION:	(47036)..(47036)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(51334)..(51334)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(51602)..(51602)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(51805)..(51805)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(55369)..(55369)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(65309)..(65309)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(65313)..(65313)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(80024)..(80024)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(100091)..(100091)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(102696)..(102696)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(105121)..(105121)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(107248)..(107248)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(117136)..(117136)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(119750)..(119750)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(119924)..(119924)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(120038)..(120038)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	NAME/key: misc.feature
LOCATION:	(121344)..(121344)
OTHER INFORMATION:	n equals a, t, g or c
FEATURE:	

```
NAME/KEY: misc.feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (152530)..(152530)
```

```
Query Match 66.7%; Score 50; DB 16; Length 1830121;
Best Local Similarity 62.2%; Pred. No. 2.1e-08;
Matches 46; Conservative 13; Mismatches 15; Indels 0; Gaps 0;
```

```
QY 2 GUGGUAUCCCAAGGAGGUAAGGACCGAUUUAUCCGAGUUCGGAUCC 61
Db 1693458 GGGGATATGCCAAGGCGTGAAGCAGCTGGTTGATCTCAGATTCTTCAATCC 1693399
QY 62 UCGUACCGCAGCCA 75
Db 1693398 TAGTACCCGAGCCA 1693385
```

```
RESULT 6
US-09-882-227-159/c
; Sequence 159, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
```

```
APPLICANT: Kleantous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
APPLICANT: Oomen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the
FILE REFERENCE: 06132/047002
CURRENT APPLICATION NUMBER: US/09/882,227
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 08/902,615
NUMBER OF SEQ ID NOS: 638
SOFTWARE: fastSeq for Windows Version 4.0
SEQ ID NO 159
LENGTH: 711
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (39)...(662)
US-09-882-227-159
```

```
Query Match 64.5%; Score 48.4; DB 10; Length 711;
Best Local Similarity 62.2%; Pred. No. 2.6e-08;
Matches 46; Conservative 12; Mismatches 16; Indels 0; Gaps 0;
```

```
QY 2 GUGGUAUCCCAAGGAGGUAAGGACCGAUUUAUCCGAGUUCGGAUCC 61
Db 82 GGGGTGTGCCAAGGCGTGAAGCAGCTGGTTGCTCTGCATTTCCAGGTTCGAATCC 23
QY 62 UCGUACCGCAGCCA 75
Db 22 TTGCACCCGAGCCA 9
```

```
RESULT 7
US-10-369-036B-3
; Sequence 3, Application US/10369036B
; Publication No. US20030228593A1
; GENERAL INFORMATION:
; APPLICANT: Suga, Hiroaki et al.
; TITLE OF INVENTION: Ribozymes with broad tRNA aminoacylation activity
; FILE REFERENCE: 11520.0290
; CURRENT APPLICATION NUMBER: US/10/369,036B
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,424
; NUMBER OF SEQ ID NOS: 61
SEQ ID NO 3
LENGTH: 74
TYPE: RNA
ORGANISM: artificial sequence
FEATURE:
NAME/KEY: misc.RNA
LOCATION:
OTHER INFORMATION: synthesized
US-10-369-036B-3
```

```
Query Match 58.4%; Score 43.8; DB 16; Length 74;
Best Local Similarity 82.7%; Pred. No. 1e-06;
Matches 62; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
```

```
QY 1 GUGGUAUCCCAAGGAGGUAAGGACCGAUUUAUCCGAGUUCGGAUCC 60
Db 1 GUGGUAUCCCAAGGAGGUAAGGACCGAUUUAUCCGAGUUCGGAUCC 59
QY 61 CUCGUACCGCAGCCA 75
Db 60 CUCGUACCGCAGCCA 74
```

RESULT 8
US-08-781-986A-4383
Sequence 4383, Application US/08781986A
Publication No. US20030054436A1
GENERAL INFORMATION:
APPLICANT: Charles Kunach
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 4383:
SEQUENCE CHARACTERISTICS:
LENGTH: 308 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-4383

Query Match 55.5%; Score 41.6; DB 8; Length 308;
Best Local Similarity 56.8%; Pred. No. 8.5e-06;
Matches 42; Conservative 11; Mismatches 21; Indels 0; Gaps 0;

QY 2 GUGUATCCCAAGGCGUAGGACCGGATUCUAUCCGAGUUCGAGUUGAUAUCC 61
DB 211 GGGCTATATNCAGCGGTAAAGCAAGGACTTGTGACTCGTCTGTTGATCC 270

QY 62 UCGUACCGCAGCCA 75
DB 271 ANCTAGCCCGACCA 284

RESULT 9
US-10-329-624-4383
Sequence 4383, Application US/10329624
Publication No. US20040043037A1
GENERAL INFORMATION:
APPLICANT: Charles Kunach
Gill H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland

COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/329,624
FILING DATE: 27-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/956,171
FILING DATE: October 20, 1997
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248PID1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 4383:
SEQUENCE CHARACTERISTICS:
LENGTH: 308 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4383:
US-10-329-624-4383

Query Match 55.5%; Score 41.6; DB 13; Length 308;
Best Local Similarity 56.8%; Pred. No. 8.5e-06;
Matches 42; Conservative 11; Mismatches 21; Indels 0; Gaps 0;

QY 2 GUGUATCCCAAGGCGUAGGACCGGATUCUAUCCGAGUUCGAGUUGAUAUCC 61
DB 211 GGGCTATATNCAGCGGTAAAGCAAGGACTTGTGACTCGTCTGTTGATCC 270

QY 62 UCGUACCGCAGCCA 75
DB 271 ANCTAGCCCGACCA 284

RESULT 10
US-09-070-927A-538/C
Sequence 538, Application US/09070927A
Patent No. US20020120116A1
GENERAL INFORMATION:
APPLICANT: Charles A. Kunach
Patrick J. Dillon
Steven Barash
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 538:
SEQUENCE CHARACTERISTICS:
LENGTH: 987 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-070-927A-538
SEQUENCE DESCRIPTION: SEQ ID NO: 538:
Query Match 54.1%; Score 40.6; DB 9; Length 987;
Best Local Similarity 57.7%; Pred. No. 2.4e-05;
Matches 41; Conservative 11; Mismatches 19; Indels 0; Gaps 0;
QY 2 GUGGUAUCCCGGAGGAGGAGCGGAUUCUAAUCCGCGAUUCCGAGUUCGAUCC 61
DB 757 GGGGTAATAGCCAGGCGTAAGGCAACGACTTGACTCCGTCATCGTGTTCGAATCC 698
QY 62 UGUACCGCAG 72
DB 697 AGCTACCCGAG 687
RESULT 11
US-08-781-986A-3701/c
Sequence 3701, Application US/08781986A
Publication No. US20030054436A1
GENERAL INFORMATION:
APPLICANT: Charles Kunach
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248BP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8512
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 3701:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-3701
Query Match 53.9%; Score 40.4; DB 8; Length 400;
Best Local Similarity 56.8%; Pred. No. 2.5e-05;
Matches 42; Conservative 11; Mismatches 21; Indels 0; Gaps 0;
QY 2 GUGGUAUCCCGGAGGAGGAGCGGAUUCUAAUCCGCGAUUCCGAGUUCGAUCC 61
DB 294 GGGCTATAGCCAGGCGTAAGGCAACGACTTGACTCCGTCATCGTGTTCGAATCC 235
QY 62 UGUACCGCAGCA 75
DB 234 AGCTACCCGAGCA 221
RESULT 12
US-10-329-624-3701/c
Sequence 3701, Application US/10329624
Publication No. US20040043037A1
GENERAL INFORMATION:
APPLICANT: Charles Kunach
Gill H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/329,624
FILING DATE: 27-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/956,171
FILING DATE: October 20, 1997
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248PID1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 3701:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3701:
US-10-329-624-3701
Query Match 53.9%; Score 40.4; DB 13; Length 400;
Best Local Similarity 56.8%; Pred. No. 2.5e-05;
Matches 42; Conservative 11; Mismatches 21; Indels 0; Gaps 0;
QY 2 GUGGUAUCCCGGAGGAGGAGCGGAUUCUAAUCCGCGAUUCCGAGUUCGAUCC 61

Db 294 GGGGTATAGCCACCGGTAAAGGCAACGACTTGTACCTCCGTCCTGTGGTTGGAATCC 235
QY 62 UCGUACCGCAGCCA 75
Db 234 AGCTAGCCCAAGCCA 221

RESULT 13
US-08-781-986A-3574
; Sequence 3574, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 3574:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1051 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-3574

Query Match 53.9%; Score 40.4; DB 8; Length 1051;
Best Local Similarity 56.8%; Pred. No. 2.9e-05;
Matches 42; Conservative 11; Mismatches 21; Indels 0; Gaps 0;

QY 2 GUGUUAUCCCCAAGGUGAAGGACCGGUAUUCUAAUCCGCAUUCGAGGUAUCC 61
Db 644 GGGGTATAGCCACCGGTAAAGGCAACGACTTGTACCTCCGTCCTGTGGTTGGAATCC 703
QY 62 UCGUACCGCAGCCA 75
Db 704 AGCTAGCCCAAGCCA 717

RESULT 14
US-10-329-624-3574
; Sequence 3574, Application US/10329624
; Publication No. US20040043037A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash

Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/329,624
; FILING DATE: 27-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/956,171
; FILING DATE: October 20, 1997
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 3574:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1051 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3574:
US-10-329-624-3574
Query Match 53.9%; Score 40.4; DB 13; Length 1051;
Best Local Similarity 56.8%; Pred. No. 2.9e-05;
Matches 42; Conservative 11; Mismatches 21; Indels 0; Gaps 0;
QY 2 GUGUUAUCCCCAAGGUGAAGGACCGGUAUUCUAAUCCGCAUUCGAGGUAUCC 61
Db 644 GGGGTATAGCCACCGGTAAAGGCAACGACTTGTACCTCCGTCCTGTGGTTGGAATCC 703
QY 62 UCGUACCGCAGCCA 75
Db 704 AGCTAGCCCAAGCCA 717
RESULT 15
US-10-398-221-6/c
; Sequence 6, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederick
; APPLICANT: GLASER, Philipp
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patentin version 3.0
; SEQ ID NO: 6
; LENGTH: 213251

TYPE: DNA
ORGANISM: Listeria innocua
US-10-398-221-6

Query Match 53.9%; Score 40.4; DB 16; Length 213251;
Best Local Similarity 55.4%; Pred. No. 6.6e-05;
Matches 41; Conservative 12; Mismatches 21; Indels 0; Gaps 0;

QY	2	GUGGUAUCCCAAGGAGGUAAGGACCGGUAUUCUAAUCCCGCAUUCGAGUUCGUAUCC	61
DB	71463	GAGCTATAGCCAAAGCGGTAAAGCAAGATTGATTCGTCATGCGCTGTTGCAATCC	71404
QY	62	UCGUACCGAGCCA	75
DB	71403	AGCTAGCCCATCTCA	71390

Search completed: October 8, 2004, 04:56:59
Job time : 207.33 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 7, 2004, 22:39:44 ; Search time 1454.67 Seconds

(without alignments)
1539.640 Million cell updates/sec

Title: US-09-721-414C-16

Perfect score: 75
Sequence: 1 gguuguaucccaaggguua.....gaauccugacagcaacca 75

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%
Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estcin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_mam:*
22: em_gss_mus:*
23: em_gss_pro:*
24: em_gss_tod:*
25: em_gss_phg:*
26: em_gss_vt1:*
27: gb_gss1:*
28: gb_gss2:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	62.8	83.7	432 12	B1941477 dg07g11.y
2	62.8	83.7	686 13	BUT24388 SUMBGF06
3	62.8	83.7	700 13	BUT24208 SUMBFI1
4	51.6	66.8	524 28	AQ990576 RfC01371

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
B1941477	432 bp	RNA	linear	EST	19-FEB-2003							
dg07g11.y1	Xenopus laevis	gastrula non normalized	Xenopus laevis									
cdna clone IMAGE:3748196	5' similar to TR:P75729	P75729	FROM BASES									
685948 TO 696574	;', mRNA sequence.											
B1941477	GI:16255949											
B1941477.1	GI:16255949											
Xenopus laevis	(African clawed frog)											
Xenopus laevis												
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;												
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;												
Xenopodine; Xenopus.												
1 (bases 1 to 432)												
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L.,												
Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y.,												
Peterson, B., Gibbons, M., Harvey, N., Rittler, E., Jackson, Y., McCann, R.,												
Waterston, R. and Wilson, R.												
Washu Xenopus EST project, 1999												
Unpublished (1999)												
Contact: Sandy Clifton, Ph.D.												
Washu Xenopus EST project, 1999												
Washington University School of Medicine												
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA												
Tel: 314 286 1800												
Fax: 314 286 1810												

ALIGNMENTS

B2009121 oe164e12.
BH493529 BOCYQ58TR
CG028902 CHGAC3YTF
BH741647 gt35902.g
CC964258 BOICT26TR
CC943811 BOIGY41TR
BH589001 BOHBF03TR
BH534699 BOCIG605TR
BH645174 BOMES07TR
BH542329 BOGGS62TF
BH720795 BOMFB21TF
BH471088 BOGIE05TR
BH741653 gt35908.g
BH470157 BOCYV86TR
CC965950 BOIEG19TF
BH501073 BOGT167TF
B2075521 1k108907.
B2009966 oe194d01.
B2490601 BOMQ147TF
BH943550 od175c12.
BH952702 od168h01.
BH669977 BOMA181TF
CC952267 BOIFJ37TR
BH482756 BOHGX85TF
BH567423 BOHKT28TR
B2013771 oe986b11.
BH501316 BOGVJ25TF
B2022595 oe990g08.
B2044419 1k158c07.
BH493017 BOHPI10TR
B2472226 BOMOE84TF
CC949117 BOICT64TR
B2058099 1k26a01.
B2437740 BOMCR15TR
BH686564 BOMNZ72TF
BH467020 BOMNR34TF
B2057550 1k755b02.
BH563424 BOGZT14TR
B2488406 BOMCG20TR
BH465538 BOCGH46TR
BH521400 BOCRC69TF

Email: eest@watson.wustl.edu
Library constructed by Bruce Blumberg
DNA sequencing by: Washington University Genome Sequencing Center
Source lab clone id - 3748196
Possible reversed clone: similarity on wrong strand This clone is available royalty-free through LBNL; contact the IMAGE Consortium (image@image.lbnl.gov) for further information.
Seq primer: -40RP from Gibco
High quality sequence stop: 409.

FEATURES

source

1. .432

Location/Qualifiers

/organism="Xenopus laevis"

/mol_type="mRNA"

/db_xref="taxon:8355"

/clone="IMAGE:3748196"

/tissue_type="gastrula (stages 10.5, 11.5 mixed)"

/lab_host="TOP-10 P"

/clone_id="Xenopus laevis gastrula non normalized"

/note="Vector: Bluescript SK-, Site 1: EcoRI; Site 2:

XhoI; cDNA was prepared from 2 ug of poly A+ RNA (equal

parts from stage 10.5 and stage 11.5 gastrulae).

EcoRI-XhoI cut cDNA was then ligated into Unizap-XR

(Stratagene) with EcoRI at the 5' end and XhoI at the 3'

end. The library was mass excised and used to infect

Top10F. Clones were picked into freezing medium (per

liter 15 g tryptone, 10g yeast extract, 5g NaCl, 36 mM

K2HPO4, 13.2 mM KH2PO4, 1.7 mM Na-citrate, 0.4 mM MgSO4 7

H2O, 6.8 mM (NH4)2SO4, 4 w/v glycerol) and grown for 24

hours. Original library construction by Bruce Blumberg

(Cho et al 1991 Cell 67, 1111-1120)."

ORIGIN

Query Match 83.7%; Score 62.8; DB 12; Length 432;

Best Local Similarity 71.6%; Pred. No. 1.5e-11;

Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GUGGUAUCCCAAGGAGGAGGACCGGAUUCUAUCCGCAUUCGAGGUAUCC 61

Db 151 GGGGTATCGCCAGCGGTAAAGCACCAGATTCTGATTCGCGCATTCGAGGTTGCAATCC 210

Qy 62 UCGUACCGCAGCCA 75

Db 211 TCGTACCCAGCCA 224

Qy 62 UCGUACCGCAGCCA 75

Db 211 TCGTACCCAGCCA 224

RESULT 2

BU724388 686 bp mRNA linear EST 23-OCT-2003

LOCUS SUMBERP1 SJM Schistosoma japonicum cDNA similar to pir|E64801

DEFINITION hypothetical protein b0663 - Escherichia coli, mRNA sequence.

ACCESSION BU724388

VERSION BU724388.1 GI:28331757

KEYWORDS EST.

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

Email: hanzg@chgc.sh.cn.

Location/Qualifiers

1. .686

/organism="Schistosoma japonicum"

/mol_type="mRNA"

/db_xref="taxon:6182"

/sex="male"

/tissue_type="whole body"

/dev_stage="adult"

/lab_host="rabbits"

/clone_id="SJM"

ORIGIN

Query Match 83.7%; Score 62.8; DB 13; Length 686;

Best Local Similarity 71.6%; Pred. No. 1.7e-11;

Matches 53; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GUGGUAUCCCAAGGAGGAGGACCGGAUUCUAUCCGCAUUCGAGGUAUCC 61

Db 144 GGGGTATCGCCAGCGGTAAAGCACCAGATTCTGATTCGCGCATTCGAGGTTGCAATCC 85

Qy 62 UCGUACCGCAGCCA 75

Db 84 TCGTACCCAGCCA 71

Qy 62 UCGUACCGCAGCCA 75

Db 84 TCGTACCCAGCCA 71

RESULT 3

BU724208 700 bp mRNA linear EST 23-OCT-2003

LOCUS SUMBERP1 SJM Schistosoma japonicum cDNA similar to pir|E64801

DEFINITION hypothetical protein b0663 - Escherichia coli, mRNA sequence.

ACCESSION BU724208

VERSION BU724208.1 GI:28331577

KEYWORDS EST.

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

ORGANISM Schistosoma japonicum

SOURCE Schistosoma japonicum

QY 62 UGUAACCGAGCCA 75
 DB 84 TCGTACCCAGCCA 71

RESULT 4
 LOCUS A0990576
 DEFINITION A0990576 524 bp DNA linear GSS 14-AUG-2000
 Rfco1371 Photorhabdus luminescens strain W14 M13 library
 Photorhabdus luminescens genomic clone PLG01371, genomic survey
 sequence.

ACCESSION A0990576
 VERSION A0990576
 KEYWORDS GI:9649170
 SOURCE GSS.
 ORGANISM Photorhabdus luminescens
 Photorhabdus luminescens
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Photorhabdus.
 1 (bases 1 to 524)

REFERENCE 1
 AUTHORS French-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T.,
 Daborn, P.J., Bowen, D. and Blattner, F.R.
 TITLE A genomic sample sequence of the entomopathogenic bacterium
 Photorhabdus luminescens W14: potential implications for virulence
 JOURNAL Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
 MEDLINE 20378633
 PUBMED 10919786
 COMMENT Contact: french-constant RH
 Department of Biology and Biochemistry
 University of Bath
 South Building, Bath BA2 7AY, UK
 Tel: (44) 1225 826621
 Fax: (44) 1225 826779
 Email: bsr@bath.ac.uk
 This is one of 2,122 random reads from the M13 library. For
 annotation of identified clones (BLASTX, BLASTN and mapping to E.
 coli K12 genome) please see french-constant et al. 2000, Nucleic
 Acids Res
 Seq primer: M13 Forward
 Classes: shotgun.

FEATURES
 source
 Location/Qualifiers
 1..524
 /organism="Photorhabdus luminescens"
 /mol_type="genomic DNA"
 /strain="W14"
 /db_xref="taxon:29488"
 /clone="PLG01371"
 /dev_stage="Primary phase variant"
 /clone_lib="Photorhabdus luminescens strain W14 M13
 library"
 /note="Genomic DNA from strain W14 was size selected (1-2
 kb) and then cloned into M13 Janus."

ORIGIN
 Query Match 68.8%; Score 51.6; DB 28; Length 524;
 Best Local Similarity 63.5%; Pred. No. 1.4e-07;
 Matches 47; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

QY 2 GUGGUAUCCCAAGGAGGUAUCCGAUUCGAGAUUCCGAGUUCGAAUCC 61
 DB 169 GGGGTATCGCAAGCGGTAAAGGACAGGTTTGGATCGGATTCGCGATTCAATCC 228
 QY 62 UGUAACCGAGCCA 75
 DB 229 TGTATCCCGAGCCA 242

RESULT 5
 LOCUS B2009121/c
 DEFINITION B2009121 732 bp DNA linear GSS 07-OCT-2002
 oel14e12.g1 B. oleracea002 Brassica oleracea genomic, genomic survey
 sequence.
 ACCESSION B2009121

VERSION B2009121.1 GI:23557538
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 732)

REFERENCE 1
 AUTHORS Delhaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
 Nash, W., Rabinowicz, P.D. and Wilson, R.K.
 TITLE Whole genome shotgun reads from Brassica oleracea
 JOURNAL Unpublished (2002)
 COMMENT Contact: Richard K. Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@watson.wustl.edu
 Plate: oel14 row: e column: 12
 Seq primer: -28RPPOT reverse
 Class: shotgun
 High quality sequence start: 105
 High quality sequence stop: 503.

FEATURES
 source
 Location/Qualifiers
 1..732
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /db_xref="taxon:3712"
 /clone_lib="B.oleracea002"
 /note="Vector: pOTW13; Whole genome shotgun library from
 flowering buds. DNA was purified from a crude nuclear
 prep using Brassica oleracea T01000D3 buds provided by
 Thomas Osborn at the University of Wisconsin. Genomic
 DNA was provided by Pablo Rabinowicz (CSHL) and the
 shotgun library prepared at Washington University Genome
 Sequencing Center."

ORIGIN
 Query Match 56.5%; Score 42.4; DB 28; Length 732;
 Best Local Similarity 58.8%; Pred. No. 0.00027;
 Matches 40; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

QY 5 GUUCCCAAGGAGGUAUCCGAUUCGAGAUUCCGAGUUCGAAUCCUG 64
 DB 153 GATATGCCAAGTGTAAAGGACCGGTTTGTGATCGGATCGCAAGGTGCAATCCGT 94
 QY 65 UACCGCAG 72
 DB 93 TACCCAG 86

RESULT 6
 LOCUS BH493529/c
 DEFINITION BH493529 198 bp DNA linear GSS 13-DEC-2001
 BOGY058TR BOGY Brassica oleracea genomic clone BOGY058, genomic
 survey sequence.
 ACCESSION BH493529
 VERSION BH493529
 KEYWORDS GI:17701633
 SOURCE GSS.
 ORGANISM Brassica oleracea
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 198)

REFERENCE 1
 AUTHORS Town, C.D., Van Aken, S., Utecher, T., Koo, H. and Fraser, C.M.
 TITLE Whole genome shotgun sequencing of Brassica oleracea
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: BOGY058TF
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends

FEATURES

source

Location/Qualifiers

1..198

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="T01000DH3"

/db_xref="taxon:3712"

/clone="BOGY58"

/clone_1lb="BOGY"

/note="Vector: PHOS1, site 1: BstXI, 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN

Query Match 54.4%; Score 40.8; DB 28; Length 198;
Best Local Similarity 57.4%; Pred. No. 0.00072;
Matches 39; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

Oy 5 GUAUCCCCAAGGGGUAAGGACCGGUAUCCGGAUCCGAGGUGCAUCCUG 64

Db 170 GTATAGCCCAAGTGTAAAGGACCGGTTTGTGTACCGGATGCAAGGTTGAAATCCTTT 111

Oy 65 UACCGCAG 72

Db 110 TACTCCAG 103

RESULT 7
CG028902 383 bp DNA linear GSS 19-AUG-2003
LOCUS CG028902

DEFINITION CHGAC37TF CHGA Cleome hassleriana genomic clone CHGAC37, genomic survey sequence.

ACCESSION CG028902.1 GI:33901058

VERSION GSS.

KEYWORDS Cleome hassleriana

SOURCE

ORGANISM Cleome hassleriana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Cleome.

1 (bases 1 to 383)

Authors: Town, C.D., Van Aken, S., Utecherack, T. and Fraser, C.M.

Title: Whole genome shotgun sequencing of Cleome hassleriana

Journal: Unpublished (2003)

Comment: Other GSS: CHGAC37TR

Contact: Chris Town

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

FEATURES

source

1..383

/organism="Cleome hassleriana"

/mol_type="genomic DNA"

/cultivar="Rose Queen"

/db_xref="taxon:28532"

/clone="CHGAC37"

/clone_1lb="CHGA"

/note="Vector: PHOS1, site 1: BstXI, 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN

Query Match 54.4%; Score 40.8; DB 29; Length 383;
Best Local Similarity 57.4%; Pred. No. 0.00084;
Matches 39; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

Oy 5 GUAUCCCCAAGGGGUAAGGACCGGUAUCCGGAUCCGAGGUGCAUCCUG 64

Db 75 GTATAGCCCAAGTGTAAAGGACCGGTTTGTGTACCGGATGCAAGGTTGAAATCCTTT 134

Oy 65 UACCGCAG 72
Db 135 TACTCCAG 142

RESULT 8

LOCUS BH741647/c

DEFINITION

g135g02.g1 Bobudo01 Brassica oleracea genomic clone g135g02 5',

genomic survey sequence.

Accession BH741647

Version BH741647.1

KeyWords GI:18976260

GSS.

Brassica oleracea

Brassicaceae

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 485)

Authors: Katari, M., O'Shaughnessy, A., Palmer, L., Bahret, A., Baker, J.,

Baile, V., Cumtius, D.M., Katzenberger, F., King, L., Kirchoff, K.,

Kult, K., Miller, B., Miller, S., Nascimben, L., Preston, R.,

Santos, L., Shah, R., Zultavern, T., Dedhia, N., Rabinowicz, P.D. and

McCombie, W.R.

Whole Genome Shotgun Reads from Brassica oleracea (2002b)

Unpublished (2002)

Contact: W. Richard McCombie

Lila Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: GC35 row: 9 column: 02

Seq primer: -21univRev

Class: Shotgun

High quality sequence stop: 485.

Location/Qualifiers

1..485

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/db_xref="taxon:3712"

/clone="g135g02"

/clone_1lb="Bobudo01"

/note="Vector: M13 for .x reads, pBluescript for .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g

reads, Site 1: EcoRV; whole genome shotgun library from .b and .g


```

VERSION      CC943811.1 GI:33814408
KEYWORDS
SOURCE       Brassica oleracea
ORGANISM     Brassica oleracea
REFERENCE    Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.
AUTHORS      Whole genome shotgun sequencing of Brassica oleracea
TITLE        Unpublished (2001)
JOURNAL      Other GSSs: BOIC26TF
COMMENT      TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA.
              Tel: 301-838-3523
              Fax: 301-838-0208
              Email: cdtown@tigr.org
              DNA is from a doubled haploid provided by Tom Osborn.
              Seg primer: TR
              Class: sheared ends.

FEATURES
  source
    Location/Qualifiers
      1..524
        /organism="Brassica oleracea"
        /mol_type="genomic DNA"
        /strain="T01000DH3"
        /db_xref="taxon:3712"
        /clone="BOIC26"
        /clone_1fb="BO_1.4.1.6_KB_nuc"
        /note="Vector: PHO52; Site 1: BstXI; 1.4-1.6 kb sheared
              nuclear DNA inserted into PHO52 using BstXI linkers"

ORIGIN
Query Match      54.4%; Score 40.8; DB 29; Length 524;
Best Local Similarity 57.4%; Pred. No. 0.0009;
Matches 39; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

QY 5 GUAUCCCAAGGAGGACCGGAAUUCUAUCCGCAUUCGAGUUGCAUCCUG 64
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 119 GTATAGCCAAAGTGTAAGGACCGGTTTGTGTAACCGCATGCAAAAGTTCCAAATCCTTT 60
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 65 UACCGCAG 72
    ||:||||
DB 59 TACTCCAG 52

RESULT 10
CC943811      552 bp      DNA      linear      GSS 18-AUG-2003
LOCUS         BOIGY41TR BO_1.4.1.6_KB_nuc Brassica oleracea genomic clone
DEFINITION    BOIGY41, genomic survey sequence.
ACCESSION     CC943811
VERSION       CC943811.1 GI:33776677
KEYWORDS      GSS.
SOURCE        Brassica oleracea
ORGANISM      Brassica oleracea
REFERENCE     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eurosid II; Brassicales; Brassicaceae; Brassica.
              1 (bases 1 to 552)
              Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.
              Whole genome shotgun sequencing of Brassica oleracea
              Unpublished (2001)
              Other_GSSs: BOIGY41TF
              Contact: Chris Town
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA.
              Tel: 301-838-3523
              Fax: 301-838-0208
              Email: cdtown@tigr.org
              DNA is from a doubled haploid provided by Tom Osborn.
              Seg primer: TR
              Class: sheared ends.

```

```

FEATURES
  source
    Location/Qualifiers
      1..552
        /organism="Brassica oleracea"
        /mol_type="genomic DNA"
        /strain="T01000DH3"
        /db_xref="taxon:3712"
        /clone="BOIGY41"
        /clone_1fb="BO_1.4.1.6_KB_nuc"
        /note="Vector: PHO52; Site 1: BstXI; 2-3 kb sheared
              nuclear DNA inserted into PHO52 using BstXI linkers"

ORIGIN
Query Match      54.4%; Score 40.8; DB 28; Length 552;
Best Local Similarity 57.4%; Pred. No. 0.00092;
Matches 39; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

QY 5 GUAUCCCAAGGAGGACCGGAAUUCUAUCCGCAUUCGAGUUGCAUCCUG 64
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 119 GTATAGCCAAAGTGTAAGGACCGGTTTGTGTAACCGCATGCAAAAGTTCCAAATCCTTT 60
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 65 UACCGCAG 72

RESULT 11
BH589001      562 bp      DNA      linear      GSS 15-DEC-2001
LOCUS         BOHBF03TR BOH8 Brassica oleracea genomic clone BOHBF03, genomic
DEFINITION    BOHBF03TR BOH8 Brassica oleracea genomic clone BOHBF03, genomic
              survey sequence.
ACCESSION     BH589001
VERSION       BH589001.1 GI:17841453
KEYWORDS      GSS.
SOURCE        Brassica oleracea
ORGANISM      Brassica oleracea
REFERENCE     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eurosid II; Brassicales; Brassicaceae; Brassica.
              1 (bases 1 to 562)
              Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.
              Whole genome shotgun sequencing of Brassica oleracea
              Unpublished (2001)
              Other_GSSs: BOHBF03TF
              Contact: Chris Town
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA.
              Tel: 301-838-3523
              Fax: 301-838-0208
              Email: cdtown@tigr.org
              DNA is from a doubled haploid provided by Tom Osborn.
              Seg primer: TR
              Class: sheared ends.

FEATURES
  source
    Location/Qualifiers
      1..562
        /organism="Brassica oleracea"
        /mol_type="genomic DNA"
        /strain="T01000DH3"
        /db_xref="taxon:3712"
        /clone="BOHBF03"
        /clone_1fb="BOH8"
        /note="Vector: PHO51; Site 1: BstXI; 2-3 kb sheared
              genomic DNA inserted into PHO51 using BstXI linkers"

ORIGIN
Query Match      54.4%; Score 40.8; DB 28; Length 562;
Best Local Similarity 57.4%; Pred. No. 0.00092;
Matches 39; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

QY 5 GUAUCCCAAGGAGGACCGGAAUUCUAUCCGCAUUCGAGUUGCAUCCUG 64
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 119 GTATAGCCAAAGTGTAAGGACCGGTTTGTGTAACCGCATGCAAAAGTTCCAAATCCTTT 60
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 65 UACCGCAG 72

```

```
Db          59 TACTCCAG 52      :|||
                                           |||
RESULT 12    BH534699        598 bp   DNA       linear   GSS 14-DEC-2001
LOCUS       BH534699
DEFINITION  BOGLG06TR BOGL Brassica oleracea genomic clone BOGLG06, genomic
              survey sequence.
ACCESSION   BH534699
VERSION     BH534699.1 GI:17763147
KEYWORDS    GSS.
SOURCE      Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 598)
AUTHORS     Town,C.D., Van Aken,S., Utepbach,T., Koo,H. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Other GSSes: BOGLG06TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel.: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source           1..598
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="T01000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGLG06"
                     /clone_1fb="BOGL"
                     /note="Vector: pHOS1, Site_1: BstXI; 2-3 kb sheared
                        genomic DNA inserted into pHOS1 using BstXI linker"
ORIGIN
Query Match      54.4%; Score 40.8; DB 28; Length 598;
Best Local Similarity 57.4%; Pred.No.0.00093;
Matches 39; Conservative 12; Mismatches 17; Indels 0; Gaps 0;
QY               65 UACGCAG 72
Db              510 GTATGCCCAAGTGTTAAGCACCGGTTTTGGTACCAGCATGCAAAAGTTGCAATCCTTT 569
                :|||
                :|||
Db              570 TACTCCAG 577
                :|||
                :|||

RESULT 13    BH645174        606 bp   DNA       linear   GSS 19-FEB-2002
LOCUS       BH645174
DEFINITION  BOMES07TR BO_2_3 KB Brassica oleracea genomic clone BOMES07,
              genomic survey sequence.
ACCESSION   BH645174
VERSION     BH645174.1 GI:18702564
KEYWORDS    GSS.
SOURCE      Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 606)
AUTHORS     Town,C.D., Van Aken,S., Utepbach,T., Koo,H. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
```

COMMENT
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
SOURCE
Location/Qualifiers
1..606
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO100DH3"
/db_xref="taxon:3712"
/clone_id="BO_2.3 KB"
/note="Vector: PHOS1; site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN

Query Match 54.4%; Score 40.8; DB 28; Length 606;
Best Local Similarity. 57.4%; Pred. No. 0.0093; Mismatches 17; Indels 0; Gaps 0;
Matches 39; Conservative 12;

DY 65 UACCGCAG 72

Dd 495 TACTCCAG 502

RESULT 14
BH542329 611 bp DNA linear GSS 14-DEC-2001
LOCUS BH542329
DEFINITION BOGS62TR BOGG Brassica oleracea genomic clone BOGS62, genomic survey sequence.
VERSION BH542329
ACCESSION BH542329.1 GI:17794110
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 611)
Town,C.D., Van Aken,S., Utechtack,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea unpublished (2001)
Other_GSSs: BOGS62TR
Contact: Chris Town

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
SOURCE
Location/Qualifiers
1..611
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO100DH3"
/db_xref="taxon:3712"
/clone_id="BOGS62"
/clone_id="BOGG"
/note="Vector: PHOS1; site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2004, 01:43:05 ; Search time 1539.33 Seconds

(without alignments)
4223.552 Million cell updates/sec

Title: US-09-721-414C-9

Perfect score: 150

Sequence: 1.ggaucgucagcagcaugaga.....gaaucucgaaaccgagcca 150

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_hcg:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	21.3	39490	3	LMFL2954
2	31.6	21.1	55713	9	AL138618
3	31.6	21.1	141605	2	AC073088
4	31.6	21.1	212105	2	AC013732
5	31.6	21.1	224044	10	AC132741
6	31.2	20.8	281747	2	AC094865
7	30.8	20.5	229649	2	AC130074
8	30.8	20.5	249337	2	AC096899
9	30.6	20.4	161593	9	HSBA49G10
10	30.4	20.3	214889	2	AC102638
11	30.4	20.3	286632	2	AC140354
12	30.2	20.1	218788	2	AC093945
13	30.2	20.1	241889	2	AC122964
14	30	20.0	110000	2	AL1954691.2
15	30	20.0	190670	2	AC122103
16	30	20.0	213097	2	AC125974
17	30	20.0	229778	2	BX005403
18	30	20.0	229609	10	AL671914
19	29.8	19.9	90463	8	AP003810
20	29.8	19.9	103024	2	AC114834.3
21	29.8	19.9	137729	10	AC079223
22	29.8	19.9	180645	10	AL733468
23	29.8	19.9	184681	2	AC084409
24	29.8	19.9	191286	2	AC112793
25	29.8	19.9	205288	2	AC140253
26	29.8	19.9	257473	2	AC132663
27	29.8	19.9	258912	2	AC129777
28	29.8	19.9	297155	2	AC095921
29	29.6	19.7	532	12	AF463831
30	29.6	19.7	153006	2	AC027782
31	29.6	19.7	218583	2	AC121165
32	29.6	19.7	218921	2	AC097912
33	29.4	19.6	141990	9	AC004691
34	29.4	19.6	154280	2	AC037480
35	29.4	19.6	180257	2	AC007044
36	29.4	19.6	214485	2	AC123416
37	29.4	19.6	215520	9	AC026956
38	29.4	19.6	268510	2	AC095840
39	29.2	19.5	3655	3	DME250955
40	29.2	19.5	67297	10	AL590876
41	29.2	19.5	165176	3	AC007588
42	29.2	19.5	202285	10	AC128024
43	29.2	19.5	229684	3	AE003815
44	29	19.3	3005	1	HVU95375
45	29	19.3	110000	2	HSY313F4_0

ALIGNMENTS

RESULT 1
LMFL2954
LOCUS LMFL2954 39490 bp DNA linear INV 19-MAR-2001
DEFINITION Leishmania major Friedlin chromosome 23 cosmid L2954.
ACCESSION AL138618
VERSION AL138618.2 GI:13397864
KEYWORDS 5S RNA; ATP-dependent DNA helicase; leu tRNA; met tRNA; PH domain
protein; possible zinc-binding transmembrane protein; prefolidin
subunit; surface antigen.
SOURCE Leishmania major
ORGANISM Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 39490)
Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and

Smith, D.F.
 TITLE A physical map of the Leishmania major Friedlin genome
 JOURNAL Genome Res. 8 (2), 135-145 (1998)
 MEDLINE 98146435
 PUBMED 9477341
 REFERENCE 2 (bases 1 to 39490)
 AUTHORS Borzym, K., Klages, S., Reinhardt, R., Beck, A., Ivens, A.C., Quail, M., Rajandream, M.A. and Barrell, B.G.
 TITLE Direct Submission
 JOURNAL Submitted (16-MAR-2001) European Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, (E-mail: barrell@sanger.ac.uk)
 D-14195 Berlin, Germany
 On Mar 21, 2001 this sequence version replaced gi:6855436.
 see http://www.ebi.ac.uk/parasites/leish.html
 Notes:
 Details of Leishmania sequencing at the Sanger Centre are available on the World Wide Web.
 see http://www.sanger.ac.uk/Projects/L_major/
 CDS are numbered using the following system eg L2954.01. L2954 (cosmid name), .01 (first CDS)
 To make the cosmid library Leishmania major Friedlin DNA was partially digested with Sau3AI prior to cloning into BamHI site of the cosmid shuttle vector cIHYG (Ryan et al. 1993 Gene 131:145-150). The sequence of the packaged vector was determined by Peter Myler and Ken Stuart at Seattle Biomedical Research Institute, and is available as accession number U59231.
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta-co is given for CDS which show significant similarity to other CDS in the database.
 Gene prediction is done using:
 (1) the FramePlot program of Bibb et al.,
 Gene 30:157-166 (1984) as implemented at http://www.nih.gov/jb-jun/cgi-bin/frameplot.pl. (2)
 codon preference based on the codon usage table for Leishmania at http://www.kazusa.or.jp/codon/
 (3) the Hexamer program which was written by Richard Durbin as an integral part of the ACEDB-based analysis tools for the C.elegans Genome Sequencing Project. The program calculates the log-likelihood score for a given DNA segment based on the frequency of 6-mers, normalised for the base-pair composition of the genome. The program was trained using a fasta file of confirmed Leishmania major coding sequences (CDS), i.e. from ATG start codon to the stop codon.
 CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg) which is preceded by a stretch of pyrimidines or part of a Kozak sequence. If this cannot be identified we choose the most upstream initiation codon. Transmembrane domains were predicted as implemented at the TMHMM server: http://www.cbs.dtu.dk/services/TMHMM-1.0/ IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid L2954 is not overlapped by a clone sequenced to date.
 Location/Qualifiers
 source 1..39490
 /organism="Leishmania major"
 /mol_type="genomic DNA"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /chromosome="23"
 /clone="cosmid L2954"
 repeat_region 168..179
 /note="(ggc)4"
 repeat_region 366..380
 /note="poly-pyrimidine tract"
 repeat_region 709..738

/note="poly-pyrimidine tract"
 repeat_region 745..759
 /note="poly-pyrimidine tract"
 repeat_region 787..798
 /note="(act)4"
 repeat_region 803..814
 /note="(gtc)4"
 repeat_region complement(913..929)
 /note="poly-pyrimidine tract"
 repeat_region 936..947
 /note="(atac)3"
 repeat_region 1024..1035
 /note="(ccc)4"
 repeat_region 1173..1184
 /note="(tcgc)3"
 repeat_region 1284..1285
 /note="(cca)4"
 gene 1405..1262
 /gene="L2954.01"
 CDS 1405..1262
 /gene="L2954.01"
 /note="L2954.01, len = 284 aa, possible surface antigen; predicted pI = 8.1969; predicted TM helices at aa 13-35 (83aa), peptide, HMW-predicted cleavage at A30-A31, 261-283; codon usage plots suggest ORF extends upstream beyond nearest stop codon, but current M1 has an excellent Kozak consensus; reasonable similarity to Q2531, surface antigen 2 (385 aa, Leishmania major, EMBL: Y09093, CAA0309); fasta scores: E(1):1.8e-22, 33.8% identity in 228 aa"
 /codon_start=1
 /label=L2954.01
 /product="possible surface antigen"
 /protein_id="CA871273.1"
 /db_xref="GI:6855437"
 /db_xref="SPTREMBL:Q9NEH5"
 /translation="MAQRARRLSNAVAVMAFVVCVLLPAGVTRALTAQNSITAEFLRSFVSYIPGASVWGDMCTWPVSCNDSPTSLITVIDNAGFTSLPELVVAGDSANIVTEIATLNDIDGGFGNMGGLKKVVLNFTNLTGTLPSVSNARSLQTVLRNSACGSLPRMTLPSTKNTDVENNLLRCSMPPTLIGLQVLSVIGNNPCCCTPSWWSRVLTSAIPQMGSAFPEPNCRPPTNCDSBAKSKSRDAPQYRDAAASCHVAVLVTVLTIVCNRSV"
 /gene="L2954.01"
 /note="predicted TM helix region, aa 13-35"
 /gene="L2954.01"
 /note="predicted TM helix region, aa 261-283"
 repeat_region 2185..2196
 /note="(ggc)4"
 repeat_region 2295..2312
 /note="poly-pyrimidine tract"
 repeat_region complement(2390..2408)
 /note="poly-pyrimidine tract"
 repeat_region 2507..2526
 /note="poly-pyrimidine tract"
 repeat_region 2589..2607
 /note="poly-pyrimidine tract"
 repeat_region 2608..2619
 /note="(acc)4"
 repeat_region 2665..2685
 /note="poly-pyrimidine tract"
 repeat_region 2691..2708
 /note="poly-pyrimidine tract"
 repeat_region 3466..3483
 /note="poly-pyrimidine tract"
 repeat_region complement(3701..3720)
 /note="poly-pyrimidine tract"
 repeat_region 4067..4085
 /note="poly-pyrimidine tract"
 gene 4766..4756
 /gene="L2954.02"
 4766..7756

```

/gene="L2954.02"
/notes="L2954.02, len = 996 aa, probably ATP-dependent DNA
helicase; predicted pi = 7.3190; contains Pfam match to
entry PF00580 UvrD-helicase, UvrD/RP helicase; contains
match to PROSITE PS00017 ATP/GTP-binding site motif A
(p-loop); reasonable similarity to several, e.g.:
REP_HAEIN, ATP-dependent DNA helicase rep (670 aa,
Haemophilus influenzae, EMBL: U32748, AAC22309); Faeta
scores: E(): 6.3e-12, 26.7% identity in 469 aa
may use M200"
/codon_start=1
/label=L2954.02
/product="probable ATP-dependent DNA helicase"
/protein_id="CA871274.1"
/db_xref="GI:6855438"
/db_xref="SPTREMBL:Q9NEH4"
/translation="MDEQIRATYAPADALILQAGSGKTQTMARITAYLQSGVP
GHSILGICFTRQAAETLRGRVSTLPTLTREAHALKTFPAHGLECRGALQAD
THVLDARQOHLARVVDYVQREKSEAVTLDVYNNVKTRMPPTIPOLDALODA
YLPYVYOKALHEHNAVDRGDIQMPYDIIRVPARALTAAGCGGGLDEQPSQEH
QPPQOQOQGWPSDVCTIALRAETHFVDFODEVEIQLALAGDACHVTCGD
PNOCIYTRGAMPNFGVWKKRPQTMALTLVNRSDGPIVEANRVKATQMAHH
REERAVTLVQASDEDELQAVPLIEHVLRRDAHLYGDIALLCSRRVQLCYVL
QSORIPVROLRGWVDHLASMSLAFRLCYSPHGEADANVTYANTAPHLSPG
AAKFFLSLSDVCOARATEARIRREMRTHVNDSEVAGDNGAQSRRSGLOPA
ATSVYGGCCPTATTPGVPFESHSPFAYIQELVYHNRSHALPKLEVKQKQNRISV
VRIYVAHAKLLAOPSCDVEQVLRVLRGGYGESMAVAANTVTSATRTKARN
SAEGWSDRCVDEASGRSSASAYQPRVGAALMERCSQNOQOORTSSAPASCSA
GGGDSRMDRHHGSRSGAASGSDGAGDEDLVLPPEEAALVMOEORNLSELVHTYHS
VOELAREVTHLEKGGGREGREASRSRSPQSGHNSRSLFPAKATPADSV
SNFASGTTTVMQALRPVAVLHRVLDYFSLVSDPSDPMIEADADAAGAA
GADTNGNGRLPTSTSPHVICGVCTYTRAKGMENPVLKSGNGEIPVAPRDEK
RIFYTMSRMAHVLCTFAABEGRGSGQAVTASVVDLPQHSOSGLTEPTYLAA
LSDTLERNVAVLDKTAIYTEQYL"
4766.6043
/misc_feature
/gene="L2954.02"
/notes="Pfam match to entry PF00580 UvrD-helicase, UvrD/RP
helicase, score 195.00, E-value 1.2e-54"
4826.4849
/misc_feature
/gene="L2954.02"
/notes="PROSITE PS00017 ATP/GTP-binding site motif A"
Query Match 21.3%; Score 32; DB 3; Length 39490;
Best Local Similarity 45.0%; Pred. No. 13;
Matches 36; Conservative 14; Mismatches 30; Indels 0; Gaps 0;
QY 27 CAGGCCCTUCACUACGUGGUCUAGGUGUAGCGGUGGCGGAGGUCUAC 86
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 23005 CATTGCTTCTACACGCGTCCCGCTAACGCTATGCGGTTCTGTGATGCTCTTC 23064
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 87 GCGUGUGUAGUCCCAAGG 106
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 23065 GGTGTGATCAGCGCAAGG 23084
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

RESULT 2
AC073088/c AC073088 55713 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-323P8 from 2, complete sequence.
ACCESSION AC073088
VERSION AC073088.5 GI:18093125
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Sultoni, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 55713)

```

```

AUTHORS Desai, A. and Abbott, A.
TITLE The sequence of Homo sapiens BAC clone RP11-323P8
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 55713)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2000) Genome Sequencing Center, Washington
University, School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 55713)
Waterston, R.
REFERENCE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 9, 2002 this sequence version replaced gi:15825633.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0323P08

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenno, M., Caranese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-207G14, 2000 bp overlap; the clone sequenced to the right is RP11-92K14, 2000 bp overlap. Actual start of this clone is at base position 67479 of RP11-207G14; actual end is at base position 13505 of RP11-92K14.

Data from AC013732 was used to finish the clone, AC073088. Polymorphisms have been identified between AC013732 AC073088.

FEATURES

```

source
1..55713
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-323P8"
/clone_11b="RPc1-11"
1..146
/repeat_family="L1"
repeat_region
152..263

```

[illegible]

Sanders, W., Saverly, G., Scherer, S., Scott, G., Shasman, S., Shen, H., Shetty, J., Snydersbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Stettin, R., Sosa, J.,

Stearle, M., Strong, R., Sutton, A., Syatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valasek, V., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

TITLE
JOURNAL
Unpublished
2 (bases 1 to 212105)
Rat Genome Sequencing Consortium.
REFERENCE
Submitted (03-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 212105)
Rat Genome Sequencing Consortium.
COMMENT
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23816487.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence only contigs. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSNH
Center clone name: CH230-355K7
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 186569 bases at least Q40
Consensus quality: 188391 bases at least Q30
Consensus quality: 189573 bases at least Q20
Estimated insert size: 197497; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_difft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 198029: contig of 198029 bp in length
* 198030 198129: gap of unknown length
* 198130 212105: contig of 13976 bp in length.
Location/Qualifiers
1. 212105
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-355K7"
3288. 4118
misc_feature
/note="Clone boundary"

clone_end:17
site:
end_sequence: B2254886"

ORIGIN
Query Match 21.1%; Score 31.6; DB 2; Length 212105;
Best Local Similarity 44.7%; Pval. No. 16;
Matches 42; Conservative 13; Mismatches 39; Indels 0; Gaps 0;

Oy 46 GUGGUAUUGGUAUUGCGGUUAGUGUGGUAUGCUAUGCGUGGUAUUGCCCAAG 105
Db 124332 GTGGGTCAAGGTTGTGCGAGGTTGGATTCATTTCACCTTTGGAGCTGCAGA 124331
Oy 106 GUACGGACCGACAUUCGAGAUUCGUAUCCUG 139
Db 124392 GTACCTGCCACACCAAGACATCGAATATACG 124425

RESULT 5
AL935326/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
MUS musculus (house mouse)
HTG.
MUS musculus (house mouse)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 224044)
Pelam, S.
Direct Submission
Submitted (08-APR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 9, 2003 this sequence version replaced gi:28971566.
Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-10117 is from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.0.
Location/Qualifiers
1. 224044
/organism="Mus musculus"

FEATURES
source

[illegible]

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
Unpublished 2 (bases 1 to 249337)	Morley K.C.	Submitted (02-Oct-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	3 (bases 1 to 249337) Rat Genome Sequencing Consortium. Direct Submission Submitted (09-May-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	On May 9, 2003 this sequence version replaced gi:24819497. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/atlas/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separate by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.	***** ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GANC Center clone name: CH230-4C18 ----- Summary Statistics Assembly program: Atlas; Consensus quality: 235927 bases at least Q40 Consensus quality: 238631 bases at least Q30 Consensus quality: 240375 bases at least Q20 Estimated insert size: 245648; sum-of-contigs estimation Quality coverage: 7x in Q20 bases; sum-of-contigs estimation ----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) * NOTE: This is a "working draft" sequence. It currently * consists of 2 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * I 247900: contig of 247900 bp in length

```

FEATURES
    * 247901 248000: gap of unknown length
    * 248001 249337: contig of 1337 bp in length.
        location/Qualifiers
            1..249337
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-4C18"
            1..1259
            /note="wgs end extension"
            clone_end:17"
            /note="wgs end extension"
            4668..6225
            /note="wgs end extension"
            clone_end:17"
            complement(7419..8227)
            /note="clone boundary"
            clone_end:17"
            site:ECORI
            end_sequence:BH304072"
            complement(244958..245819)
            /note="clone boundary"
            clone_end:Sp6
            site:ECORI
            end_sequence:BH304074"

ORIGIN
Query Match      20.5%  Score 30.8; DB 2; Length 249337;
Best Local Similarity 40.8%; Pred. No. 29;
Matches 40; Conservative 16; Mismatches 42; Indels 0; Gaps 0;

Qy      34  UUCUCACUACAGGUGGUCAGUAGGCGGUAAGGCGGAGUACUACGCGUG 93
       :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      34625 TTCTGACTGACTTGATGCTGATGAGGCTGATGAGCTGATGCTGAAATC 34566
       :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Qy      94  GUAUCCCAAGGAGGAGCAGCAGCAGACUUCAGAUUC 131
       :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      34565 CTCCTCCGAGGACAGGCTGATGAGAGAAATTCG 34528
       :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 9
HSB499310/c      161593 bp      DNA      linear      PRI 20-JUN-2001
LOCUS
DEFINITION
Human DNA sequence from clone RP11-49G10 on chromosome 20 Contains
a gene similar to bovine salivary protein BSP30, the LOC51297 gene
for LUNX protein, 5' end of a gene encoding a protein similar to
murine von Ebner minor salivary gland protein, a novel gene, a
pseudogene similar to STAT-induced STAT inhibitor-2, a pseudogene
similar to ribosomal protein L12, a putative novel transcript,
ESTs, STSs, GSSs and a Cpg Island, complete sequence.
AL121901
AL121901.20 GI:8249854
HTG: BSP30; Cpg Island; LOC51297; LUNX; RPL12; STAT-inhibitor.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 161593)
Tracey A.
Direct Submission
Submitted (14-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requesters: clonequery@sanger.ac.uk
On Jun 5, 2000 this sequence version replaced gi:8017404.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at

```

```

FEATURES
    source
        http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
        was generated from part of bacterial clone contigs of human
        chromosome 20, constructed by the Sanger Centre Chromosome 20
        Mapping Group. Further information can be found at
        http://www.sanger.ac.uk/HGP/Chr20
        IMPORTANT: This sequence is not the entire insert of clone
        RP11-49G10 it may be shorter because we sequence overlapping
        sections only once, except for a 100 base overlap.
        The true left end of clone RP11-49G10 is at 1 in this sequence.
        The true left end of clone RP5-118704 is at 161494 in this sequence.
        The true right end of clone RP4-733023 is at 27823 in this
        sequence. This sequence was finished as follows unless otherwise
        noted: all regions were either double-stranded or sequenced with an
        alternate chemistry or covered by high quality data (i.e., phred
        quality >= 30); an attempt was made to resolve all sequencing
        problems, such as compressions and repeats; all regions were
        covered by at least one plasmid subclone or more than one M13
        subclone; and the assembly was confirmed by restriction digest.
        RP11-49G10 is from the library RPCI-11.1 constructed by the group
        of Pieter de Jong. For further details see
        http://www.chori.org/bacpac/home.htm
        VECTOR: pBACE3.6.
        Location/Qualifiers
            1..161593
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="20"
            /clone="RP11-49G10"
            /clone_1lb="RPCI-11.1"
            complement(1..107)
            /note="match: GSS: Em:AQ067627"
            6..706
            /note="match: GSS: Em:AQ051854"
            137..548
            /note="match: GSS: Em:AQ050113"
            530..603
            /note="37 copies 2 mer ca 66% conserved"
            533..686
            /note="2 copies 77 mer 83% conserved"
            613..678
            /note="33 copies 2 mer ca 75% conserved"
            3030..3174
            /note="L1B3 repeat: matches 5941..6084 of consensus"
            5140..5183
            /note="22 copies 2 mer ct 75% conserved"
            5184..5344
            /note="L1A2 repeat: matches 5986..6146 of consensus"
            complement(6104..6463)
            /note="match: GSS: Em:AQ39300"
            6120..6239
            /note="2 copies 60 mer 85% conserved"
            complement(8476..9002)
            /note="match: GSS: Em:AQ537190"
            complement(14766..15061)
            /note="match: STS: Em:HS922114S"
            14772..14779
            /note="Random repeat. Forced join. Approximately 500 bases
            missing according to restriction digest."
            16219..16387
            /note="match: GSS: Em:AL001671"
            16225..16387
            /note="match: GSS: Em:AL013146"
            18325..18826
            /note="match: GSS: Em:AQ56113"
            18328..18828
            /note="match: GSS: Em:AQ532957"
            18333..18770
            /note="match: GSS: Em:AQ765930"
            complement(20975..21245)
            /note="match: GSS: Em:AZ102442"
            20999..21238
            /note="4 copies 60 mer 69% conserved"

```

misc_feature	21031..21263 /note="match: GSS: Em:A2055779 Em:A2084220"
repeat_region	21144..21225 /note="41 copies 2 mer ag 75% conserved"
misc_feature	/note="match: GSS: Em:A0395327" complement(21172..21428)
misc_feature	/note="match: GSS: Em:AQ727877" complement(23473..23950)
misc_feature	/note="match: GSS: Em:A0691469" complement(23637..23948)
misc_feature	/note="match: GSS: Em:AQ152593" 23950..24355
misc_feature	/note="match: GSS: Em:AQ508832" 23955..24456
misc_feature	/note="match: GSS: Em:A0636646" 26979..27422
repeat_region	/note="match: GSS: Em:A0636646" 31882..32061
repeat_region	/note="3 copies 60 mer 73% conserved" 31921..32032
repeat_region	/note="56 copies 2 mer ct 63% conserved" 33693..33762
repeat_region	/note="MER47 repeat: matches 2251..2322 of consensus" 33733..33785
gene	/note="MER47 repeat: matches 1..55 of consensus" 35535..48787
mRNA	/gene="bA49G10.1" join(35555)..35585,36507..36678,40308..40452,41455..41562, 42783..42935,45523..45604,46980..47043,47896..47973, 48600..48787) /gene="bA49G10.1" /product="bA49G10.1 (similar to bovine salivary protein BSP30)" /note="match: cDNAs: Em:U79414 Em:U79413 Em:AX061621 match: ESTs: Em:A1C54622" /evidence=not experimental join(36522..36678,40308..40452,41455..41562,42783..42935, 45523..45604,46980..47043,47896..47973) /gene="bA49G10.1" /note="match: proteins: Tr:P79125 Sw:P07743" /codon start=1 /evidence=not experimental /product="bA49G10.1 (similar to bovine salivary protein BSP30)" /protein_id="CAC03546.1" /db_xref="GI:9801234" /db_xref="SPTREMBL:O9B000" translation="MQLMKVILCGVLTCGTSESLLDNLGNDSNVNDEKPYHIEGL EVDNTLKIKELKYDGCYLQKSSAMQAKQAERKLNNVISKLPITNDIFGL KISNSLIDLVKAEPIDDGKGLNLSFPYNAVAPPIIGIINLKASIDLTVATIER DQIQPVAVLGECASDPSISLSDLRHSQTINFVSINTLSTYSLSLOKEICP LIRFIHSLDVAVIIQQVNDNPQHKTQLTL" 38155..38751 /note="LTPA8 repeat: matches 4514..5125 of consensus" 38750..39785 /note="LTPA8 repeat: matches 5128..6163 of consensus" 48769..48774 /gene="bA49G10.1" 48787 /gene="bA49G10.1" 55540..53450 /note="CPG island" /evidence=not experimental 53831..54141 /note="match: STS: Em:G31731" complement(53831..54131) /note="match: GSS: Em:AQ897745" 53842..53969 /note="match: GSS: Em:AQ429347" complement(53844..54141) /note="match: GSS: Em:AF101963" 53858..54022 /note="match: GSS: Em:AQ392891" complement(53908..54141)

Query Match	20.4%	Score 30.6	DB 9	Length 161593
Best Local Similarity	51.9%	Pred. No. 35		
Matches	40	Conservative	8	Mismatches 29; Indels 0; Gaps 0;
Qy	40	CUAAACGGUGGUCUAGUGGUAUUGGCGUAGUGCGGGAUUCUACUACGUGGUGGUAUCC	99	/note="match: GSS: Em:AQ415287" 53958. .54139 /note="match: GSS: Em:AQ378081" 53970. .54139 /note="match: GSS: Em:AQ527291" 53972. .54142 /note="match: GSS: Em:AQ429347" 54000. .54141 /note="match: GSS: Em:AQ392510" complement(54027. .54139) /note="match: GSS: Em:AQ533658" 54030. .54138 /note="match: GSS: Em:AG000935" 54041. .54149 /note="match: GSS: Em:AQ392891" 54767. .55668 /gene="bA49G10.2" /pseudo 54767. .55668 /gene="bA49G10.2"
Db	34523	CAAAACGGTGGGACAAACACAGTGGCTCCAGAGCTGCTTTCTTACACTGACCATGGCT	34466	/note="bA49G10.2 (similar to STAT-induced STAT inhibitor-2)
Qy	100	CCAAAGCGUACGGACCG	116	
Db	34463	CCACTGTTCAGACACAG	34447	
RESULT 10				
AC102638		214889 bp	DNA	linear
LOCUS				HTG 11-DEC-2003
DEFINITION		Mus musculus chromosome 3 clone RP23-406D2 map 3, *** SEQUENCING IN PROGRESS ***.		
ACCESSION		AC102638		
VERSION		AC102638.11		GI:39725798
KEYWORDS		HTG; HTGS PHASE2; HTGS_FULFILLTOP; HTGS_ACTIVEFIN.		
SOURCE		Mus musculus (house mouse)		
ORGANISM		Mus musculus		
REFERENCE		Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS		1 (bases 1 to 214889)		
TITLE		Bitren,B., Nusbaum,C. and Lander,E.		
JOURNAL		Mus musculus chromosome 3, clone RP23-406D2		
REFERENCES		Unpublished		
AUTHORS		2 (bases 1 to 214889)		
TITLE		Bitren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barta,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campioni,A., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., DeArillano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,D., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Hatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McKean,P., McKernan,K., McPheters,R., Meldrum,J., Meneses,L., Mihova,T., Miengo,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Riee,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thumann,N., Stojanovic,N., Stranas,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliev,H.,		

Viell, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Submitted (23-NOV-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 214889)
 Birtten, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, V. S., Dodge, S., Dooley, K., Dorris, L., Erickson, V., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karateas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melidrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunthang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viell, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (11-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Dec 11, 2003 this sequence version replaced gi:38564380.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center -----
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WISR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information -----
 Center project name: 119145
 Center clone name: 406_D_2

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 214889: contig of 214889 bp in length.
 Location/Qualifiers
 1. 214889
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="3"
 /map="3"
 /clone="RP23-406D2"
 /clone_11b="RP23 Female Mouse BAC"
 ORIGIN
 Query Match 20.3% Score 30.4; DB 2; Length 214889;
 Best Local Similarity 44.3%; Pred. No. 40; Indels 0; Gaps 0;
 Matches 39; Conservative 13; Mismatches 36;
 46 GUGGUCACUAGGUAUUGCCGUAUGGCGGAGUACUACUGGUGUAUCCCAAGG 105
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 93205 GTGGCTCAAGGTTTGTGACAGCGGTGTGCTGCTCATTTCTCTGTGTGACCTAGGA 93264
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Oy 106 GUACGGACCGGACAUUCGAGAUUGCAA 133
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Db 93265 GTACCTGCTGCACCAAGAGACTAGAA 93292
 RESULT 11
 AC140354
 LOCUS
 DEFINITION
 Mus musculus chromosome UNK clone RP24-249A13, WORKING DRAFT
 SEQUENCE, 13 unordered pieces.
 AC140354
 AC140354.1 GI:28475611
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 286632)
 McPherson, J.D. and Waterston, R.H.
 The sequence of Mus musculus clone
 Unpublished
 2 (bases 1 to 286632)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submissions@wustl.wustl.edu
 Project Information
 Center project name: M.BB0249A13
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: Plasmid; 100%
 Chemistry: Dye-Primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.99019
 Consensus quality: 281709 bases at least Q40
 Consensus quality: 282799 bases at least Q30
 Consensus quality: 283475 bases at least Q20

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1382: contig of 1382 bp in length
 1382: gap of unknown length
 1483: gap of 3838 bp in length
 5320: gap of unknown length
 5420: gap of 3297 bp in length
 8717: gap of unknown length
 8817: gap of 5398 bp in length
 14215: contig of 5398 bp in length
 14315: gap of unknown length
 14316: gap of 3681 bp in length
 17996: contig of 17996 bp in length
 18096: gap of unknown length
 18097: gap of 6074 bp in length
 24171: gap of unknown length
 24270: gap of 19803 bp in length
 44073: contig of 19803 bp in length
 44074: gap of unknown length
 44174: gap of 2117 bp in length
 65350: contig of 2117 bp in length
 65351: gap of unknown length
 65450: gap of 15059 bp in length
 80509: contig of 15059 bp in length
 80609: gap of unknown length
 88761: contig of 18152 bp in length
 98762: gap of unknown length
 98861: gap of unknown length
 98862: gap of 28066 bp in length
 126929: contig of 28066 bp in length
 126930: gap of unknown length

FEATURES
* 127030 159839: config of 32810 bp in length
* 159840 159939: gap of unknown length
* 159940 286632: config of 126693 bp in length
Location/Qualifiers

ORIGIN

Query Match	20.3%	Score 30.4;	DB 2;	Length 286632;
Best Local Similarity	45.0%;	Pred. No. 39;		
Matches 36;	Conservative 13;	Mismatches 31;	Indels 0;	Gaps 0;

OY		38	CACUACGGUGGAGUCAGUAUUUGGUGAGUGCCGGAUUGUACUACGUGUGUGUUV	97
			: : : : : : : : : : : : :	
Dd		66544	CCCATATCTCTGCAGGTTTTTCGCAATTATGTGCCCTTCTCCTCGACTAT	66603
OY		98	CCCCAAGGUAACGGGACCGG	117
Dd		66604	CCTCATGGGGCTTGACCTG	66623

RESULT 12	AC093945/c	LOCUS	AC093945	218788 bp	DNA	linear	HTG 10-MAY-2003
DEFINITION	Rattus norvegicus clone CH230-14303, WORKING DRAFT SEQUENCE.						
ACCESSION	AC093945						
VERSION	AC093945.7	GI:30520594					
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.						
SOURCE	Rattus norvegicus (Norway rat)						
ORGANISM	Rattus norvegicus						
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus;						
REFERENCE	1 (bases 1 to 218788)						
AUTHORS	Mundy D,Marle., Metzker M,lee., Abramzon,S., Adams,C., Alder,J.,						

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (baaes 1 to 218788)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (12-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (baaes 1 to 218788)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On May 10, 2003 this sequence version replaced gi:24942675.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
 Hollins,B., Howells,S., Hulak,S., Hume,J., Idelbrd,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kover,C.,
 Kows,C., Kratt,C.L., Lebow,H., Leyan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorenshueber,L., Louieged,H., Lozada,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindratne,M., Mahmod,R., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mawmney,S., McLeod,M.P., McKenell,T.Z., Neenan,E.,
 Milosavljevic,A., Miner,G., Ming,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Muntalasa,M., Murphy,M., Nair,L.,
 Nankervyls,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
 Nwackelmech,O., Okunnu,G., Olarnpinasoon,A., Pal,S., Parks,K.,
 Pasernak,S., Paul,H., Perez,A., Perez,L., Frankoch,C.,
 Plopper,F., Polndexter,A., Popovic,C.D., Primus,E., Pu,L.,L.,
 Puaao,M., Quitorz,T., Rachtin,E., Reeves,K., Regier,M., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shaltsman,S., Shen,H.,
 Shetty,J., Shvarcbeyn,A., Sison,I., Sitter,C.D., Smay,D.,
 Sneed,A., Sodegrien,E., Song,X.-Z., Sorrell,R., Soes,J.,
 Steimle,M., Strong,R., Sulton,A., Svacek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
 Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,D.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wlaczek,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,Y., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
 Niederhuesern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinsteht,G. and Gibbs,R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 241889)
 Worley,K.C.
 Direct Submission
 Submitted (26-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 241889)
 Rat Genome Sequencing Consortium.
 Submitted (12-SEP-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Sep 19, 2002 this sequence version replaced gi:21908810.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the
 sequence may extend beyond the ends of the clone and there may be
 contigs that consist entirely of whole genome shotgun sequence
 reads. Both end sequences and whole genome shotgun sequence only
 contigs will be indicated in the feature table.

 Genome Center

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information

 Center project name: GYBK
 Center clone name: CH230-33K9

 Summary Statistics
 Assembly program: Phrap, version 0.990329
 Consensus quality: 225569 bases at least Q40
 Consensus quality: 228593 bases at least Q40
 Consensus quality: 230509 bases at least Q20
 Estimated insert size: 243943, sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases, sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a "working draft" sequence. It currently
 consists of 2 contigs. The true order of the pieces


```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXPX
Center clone name: CH230-430024
----- Summary Statistics
Assembly program: Phrap, version 0.990129
Consensus quality: 178412 bases at least Q40
Consensus quality: 179811 bases at least Q30
Consensus quality: 180822 bases at least Q20
Estimated insert size: 183611; sum-of-coverage estimation
Quality coverage: 7x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 187191: contig of 187191 bp in length
* 187192 187291: gap of unknown length
* 187292 190670: contig of 3379 bp in length.
Location/Qualifiers
1.190670
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-430024"
1.1377
/note="wgs_contig"
179368.180060
/note="clone boundary
clone_end:5p6
site:
end_sequence:RXBN96TVB"
181446.183442
/note="wgs_end_extension
clone_end:5p6"
185607.187191
/note="wgs_end_extension
clone_end:5p6"
187292.188701
/note="wgs_end_extension
clone_end:5p6"

misc_feature
misc_feature
misc_feature
misc_feature

ORIGIN

Query Match 20.0%; Score 30; DB 2; Length 190670;
Best Local Similarity 37.3%; Pred.No. 54;
Matches 38; Conservative 19; Mismatches 45; Indels 0; Gaps 0;

OY 38 CACUACGUGGUGGUAUUGCGUUNAGUGCGGAGUCUACUACGUGGUGUAU 97
Db 96642 CACTTACTGTAGTCAAGGTTTGTGCTGGCTGCTTATGTTCTTTGGTAT 96583
OY 98 CCCAAGGUGGAGCCGACAUUCGAGAUUCGACUCCUG 139
Db 96582 CCTCGGAAATCCTTCTGCAAAAGACACTAGAGTTGG 96541

```

Search completed: October 8, 2004, 04:44:46
 Job time : 1549.33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 7, 2004, 22:35:09 : Search time 383.333 Seconds
(without alignment) 1662.338 Million cell updates/sec

Title: US-09-721-414c-9

Perfect score: 150

Sequence: 1 ggaucgucgucgucgucgagca.....gaucgucgucgucgagca 150

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29yand04:*
1: geneseqn19808:*
2: geneseqn19908:*
3: geneseqn20008:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	150	100.0	150	4 AAC85925	AAC85925 Pre-29 ca
2	150	100.0	150	4 AAC85927	AAC85927 Pre-24 ca
3	136.4	90.9	149	4 AAC85928	AAC85928 Pre-25 ca
4	136.4	90.9	149	4 AAC85931	AAC85931 Pre-19 ca
5	135.4	90.3	149	4 AAC85930	AAC85930 Pre-5 cat
6	134.8	89.9	149	4 AAC85929	AAC85929 Pre-22 ca
7	127	84.7	150	4 AAC85926	AAC85926 Pre-36 ca
8	126	84.0	148	4 AAC85933	AAC85933 Pre-23 ca
9	124.4	82.9	150	4 AAC85932	AAC85932 Pre-12 cat
10	122	81.3	146	4 AAC85923	AAC85923 Pre-38 ca
11	115.6	77.1	151	4 AAC85924	AAC85924 Pre-12 ca
12	70.6	47.1	160	4 AAC85937	AAC85937 H2 Leu ca
13	39	26.0	73	4 AAC85934	AAC85934 V1 varian
14	38	25.3	74	9 ADCl6643	ADCl6643 Aminoacyl
15	36.4	24.3	156	4 AAC85938	AAC85938 DI-Leu ca
16	35.8	23.9	73	4 AAC85936	AAC85936 V3 varian
17	35.4	23.6	56	9 ADCl6652	ADCl6652 Aminoacyl
18	33	22.0	101	9 ADCl6648	ADCl6648 Aminoacyl
19	30.6	20.4	75	4 AAC85935	AAC85935 V2 varian
20	29.8	19.9	77	9 ADCl6686	ADCl6686 Aminoacyl
21	29.4	19.6	57	9 ADCl6641	ADCl6641 Aminoacyl
22	29.2	19.5	76	9 ADCl6642	ADCl6642 Aminoacyl
23	29.2	19.5	3658	4 AB114055	AB114055 Drosophil

C	24	29.2	19.5	13033	4	AB114054	AB114054 Drosophil
C	25	29	19.3	75	4	AAC85943	AAC85943 ccrRNA, 8/
C	26	29	19.3	76	4	AAC85940	AAC85940 ccrRNA, DNA
C	27	29	19.3	38272	4	AAC84952	AAC84952 Human imm
C	28	29	19.3	38348	4	AAC84953	AAC84953 Human imm
C	29	28.8	19.2	1990	4	AA160655	AA160655 Human pol
C	30	28.8	19.2	2050	8	AA158869	AA158869 Human pol
C	31	28.8	19.2	2050	8	AD548851	AD548851 Novel hum
C	32	28.6	19.1	35	9	ADCl6689	ADCl6689 Aminoacyl
C	33	28.6	19.1	678	2	AAx08416	AAx08416 5' flank
C	34	28.6	19.1	1078	7	ABT18585	ABT18585 Aspergill
C	35	28.6	19.1	1078	7	ABT20401	ABT20401 Aspergill
C	36	28.6	19.1	3077	7	ABT17991	ABT17991 Aspergill
C	37	28.6	19.1	3078	7	ABT19805	ABT19805 Aspergill
C	38	28.4	18.9	3292	5	ABAI6966	ABAI6966 Human ner
C	39	27.4	18.3	100	7	ACD68841	ACD68841 E. coli K
C	40	27.4	18.3	644	6	AA147112	AA147112 Peptide p
C	41	27.4	18.3	2007	5	AA694249	AA694249 DNA encod
C	42	27.4	18.3	2383	5	AA679656	AA679656 DNA encod
C	43	27.4	18.3	3175	5	AA687641	AA687641 DNA encod
C	44	27.4	18.3	3175	5	AA692254	AA692254 DNA encod
C	45	27.2	18.1	362	5	AA64871	AA64871 Novel hum

ALIGNMENTS

RESULT 1	AAC85925	AAC85925 standard; RNA; 150 BP.
AC	AAC85925;	
XX		
AC	AAC85925;	
XX		
DT	22-AUG-2001 (first entry)	
XX		
DE	Pre-29 catalytic RNA.	
XX		
KW	Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;	
KW	amino acid; aminoacylation; ss.	
XX		
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	misc_RNA	1..89
FT		/*tag= a
FT		/label= 5' leader domain
FT	crRNA	90..150
FT		/*tag= b
FT		/label= tRNA domain
FT	stem_loop	99..113
FT		/*tag= c
FT		/label= D-loop
FT	D_loop	120..123
FT		/*tag= d
FT		/label= V-loop
FT	stem_loop	124..140
FT		/*tag= e
FT		/label= T-loop
PN	MO200138582-A1.	
PD	31-MAY-2001.	
XX		
PF	22-NOV-2000; 2000MO-US032184.	
XX		
PR	24-NOV-1999; 99US-0167331P.	
XX		
PR	28-JUN-2000; 2000US-0214382P.	
XX		
PA	(UNY) UNIV NEW YORK STATE RES FOUND.	
XX		
PI	Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzaghi N;	
XX		
DR	WPI; 2001-381295/40.	

XX Catalytic RNA having aminoacylating activity comprises a tRNA-like and a
PT ribozyme domain and is useful to produce non-natural amino acids for
PT biomedical and therapeutic purposes.

XX Claim 1; Page 32; 38pp; English.

XX The sequences given in AAC85923-36 are catalytic RNA molecules,
CC designated cis-aminoacylating RNA, which comprise a tRNA-like and a
CC ribozyme domain, and which have the ability to selectively amino-acylate
CC their own 3' terminus with specific amino acids. The ribozyme domain has
CC the catalytic activity and also confers amino acid specificity for
CC aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
CC -like molecules in cis or in trans for the production of proteins
CC containing non-natural amino acids that may be of use for biomedical or
CC therapeutic purposes. The catalytic RNA molecules were constructed by
CC applying selection to a randomly synthesised RNA pool. By attaching the
CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
CC contacting the complex with a substrate molecule (natural or non-natural
CC amino acid), self- aminoacylating RNA molecules were identified

XX Sequence 150 BP; 29 A; 38 C; 47 G; 0 T; 36 U; 0 Other;

XX Query Match 100.0%; Score 150; DB 4; Length 150;

XX Best Local Similarity 100.0%; Pred. No. 9.1e-45; Mismatches 0; Indels 0; Gaps 0;

XX Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAUCGUAUGCAUUGAGAUUUCGCGAGCCUUCUACUAACGGUGGUCUAUGGUAU 60

Db 1 GGAUCGUAUGCAUUGAGAUUUCGCGAGCCUUCUACUAACGGUGGUCUAUGGUAU 60

QY 61 UGGCGUUAUGGUGCGGAGUACUAACGUGUGUAUCCCAAGGAGUACGGAGCCGAGCA 120

Db 61 UGGCGUUAUGGUGCGGAGUACUAACGUGUGUAUCCCAAGGAGUACGGAGCCGAGCA 120

QY 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

XX 31-MAY-2001.

XX 22-NOV-2000; 2000MO-US032184.

XX 24-NOV-1999; 99US-0167331P.

XX 28-JUN-2000; 2000US-0214382P.

XX (UNYNY) UNIV NEW YORK STATE RES FOUND.

XX Hitoaki S, Dimitrios K, Hirohide S, Lee N, Bonzaghi N;

XX WPI; 2001-381295/40.

XX Catalytic RNA having aminoacylating activity comprises a tRNA-like and a

XX ribozyme domain and is useful to produce non-natural amino acids for

XX biomedical and therapeutic purposes.

XX Claim 1; Page 33; 38pp; English.

XX The sequences given in AAC85923-36 are catalytic RNA molecules,
CC designated cis-aminoacylating RNA, which comprise a tRNA-like and a
CC ribozyme domain, and which have the ability to selectively amino-acylate
CC their own 3' terminus with specific amino acids. The ribozyme domain has
CC the catalytic activity and also confers amino acid specificity for
CC aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
CC -like molecules in cis or in trans for the production of proteins
CC containing non-natural amino acids that may be of use for biomedical or
CC therapeutic purposes. The catalytic RNA molecules were constructed by
CC applying selection to a randomly synthesised RNA pool. By attaching the
CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
CC contacting the complex with a substrate molecule (natural or non-natural
CC amino acid), self- aminoacylating RNA molecules were identified

XX Sequence 150 BP; 29 A; 38 C; 47 G; 0 T; 36 U; 0 Other;

XX Query Match 100.0%; Score 150; DB 4; Length 150;

XX Best Local Similarity 100.0%; Pred. No. 9.1e-45; Mismatches 0; Indels 0; Gaps 0;

XX Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAUCGUAUGCAUUGAGAUUUCGCGAGCCUUCUACUAACGGUGGUCUAUGGUAU 60

Db 1 GGAUCGUAUGCAUUGAGAUUUCGCGAGCCUUCUACUAACGGUGGUCUAUGGUAU 60

QY 61 UGGCGUUAUGGUGCGGAGUACUAACGUGUGUAUCCCAAGGAGUACGGAGCCGAGCA 120

Db 61 UGGCGUUAUGGUGCGGAGUACUAACGUGUGUAUCCCAAGGAGUACGGAGCCGAGCA 120

QY 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

Db 121 UUCGAGAUUUGCAUUCUCCGUAACCGCAGCCA 150

XX RESULT 3

XX AAC85928 standard; RNA; 149 BP.

XX AAC85928;

XX 22-AUG-2001 (first entry)

XX Pre-25 catalytic RNA.

XX Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;

XX amino acid; aminoacylation; ss.

XX Synthetic.

XX Key misc_RNA

XX Location/Qualifiers

XX /*tag= a

XX /label= 5' leader domain

XX tRNA 89.149

FT	/*tag= b	FT	/label= tRNA domain
FT	98. .111	FT	stem_loop
FT	/*tag= c	FT	/label= D-loop
FT	118. .121	FT	/*tag= d
FT	/label= V-loop	FT	stem_loop
FT	122. .138	FT	/*tag= e
FT	/label= T-loop	FT	
PN	MO200138582-A1.	PN	
PD		PD	
XX	31-MAY-2001.	XX	
PF	22-NOV-2000; 2000MO-US032184.	PF	
PR	24-NOV-1999; 99US-0167331P.	PR	
PR	28-JUN-2000; 2000US-0214382P.	PR	
XX	(UNYNY) UNIV NEW YORK STATE RES FOUND.	XX	
PA		PA	
XX	Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzagni N;	XX	
PI	WPI; 2001-381295/40.	PI	
XX		XX	
DR		DR	
XX		XX	
PT	Catalytic RNA having aminoacylating activity comprises a tRNA-like and a	PT	
PT	ribozyme domain and is useful to produce non-natural amino acids for	PT	
XX	biomedical and therapeutic purposes.	XX	
PS	Claim 1; Page 33; 38pp; English.	PS	
CC	The sequences given in AAC85923-36 are catalytic RNA molecules,	CC	
CC	designated c1e-aminoacylating RNA, which comprise a tRNA-like and a	CC	
CC	ribozyme domain, and which have the ability to selectively amino- acylate	CC	
CC	their own 3' terminus with specific amino acids. The ribozyme domain has	CC	
CC	the catalytic activity and also confers amino acid specificity for	CC	
CC	aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA	CC	
CC	-like molecules in cis or in trans for the production of proteins	CC	
CC	containing non-natural amino acids that may be of use for biomedical or	CC	
CC	therapeutic purposes. The catalytic RNA molecules were constructed by	CC	
CC	applying selection to a randomly synthesised RNA pool. By attaching the	CC	
CC	RNA molecules of this pool to the 5' end of a tRNA-like molecule and	CC	
CC	contacting the complex with a substrate molecule (natural or non-natural	CC	
CC	amino acid), self- aminoacylating RNA molecules were identified	CC	
XX		XX	
XX	Sequence 149 BP; 30 A; 38 C; 45 G; 0 T; 36 U; 0 Other;	XX	
SO		SO	
Query Match	90.9%; Score 136.4; DB 4; Length 149;	Query Match	
Best Local Similarity	98.7%; Pred. No. 8,6e-40;	Best Local Similarity	
Matches 148; Conservative	0; Mismatches 1; Indels 1; Gaps 1	Matches 148; Conservative	
OY	1 GGAUGUGCAUGUCAUUGAGAUUUCGCGAGCCCTUUCUACUAACGGUGGUCUUGGGUUAU 60	OY	
DB	1 GAUUGUCAUGUCAUUGAGAUUUCGCGA-GCCUUCUACUAACGGUGGUCUUGGGUUAU 59	DB	
OY	61 UGGCGUUAUGUGCGGGAUGUCUAUCGUGUGGUAUUCGCCAAGGGUACGGAGCGGGACA 120	OY	
DB	60 UGGCGUUAUGUGCGGGAUUAUCUAUCGUGUGGUAUUCGCCAAGGGUACGGAGCGGGACA 119	DB	
OY	121 UUCGAGAUUUGCAUUCUCUGUACCGGACGCCA 150	OY	
DB	120 UUCGAGAUUUGCAUUCUCUGUACCGGACGCCA 149	DB	
RESULT 4		RESULT 4	
ID	AAC85931 standard; RNA; 149 BP.	ID	
XX	AAC85931;	XX	
XX		XX	
DT	22-AUG-2001 (first entry)	DT	

XX	DE	Pre-19 catalytic RNA.
XX	XX	Catalytic RNA; cis-aminocyclating RNA; tRNA-like domain; ribozyme domain;
KW	KW	amino acid; aminoacylation; ss.
XX	XX	Synthetic.
OS	OS	
XX	XX	
FT	FT	Key
FT	FT	misc_RNA
FT	FT	Location/Qualifiers
FT	FT	1..88
FT	FT	/tag= a
FT	FT	/label= 5', leader domain
FT	FT	89..149
FT	FT	/tag= b
FT	FT	/label= tRNA domain
FT	FT	98..111
FT	FT	/tag= c
FT	FT	/label= D-loop
FT	FT	118..121
FT	FT	/tag= d
FT	FT	/label= V-loop
FT	FT	122..138
FT	FT	/tag= e
FT	FT	/label= T-loop
XX	XX	
PN	PN	WO200138582-A1.
PD	PD	
XX	XX	31-MAY-2001.
PE	PE	22-NOV-2000; 2000WO-US032184.
XX	XX	
XX	XX	24-NOV-1999; 99US-0167331P.
PR	PR	28-JUN-2000; 2000US-0214382P.
XX	XX	
PA	PA	(UNYNY) UNIV NEW YORK STATE RES FOUND.
XX	XX	
PI	PI	Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzagni N;
XX	XX	WP1; 2001-381295/40.
XX	XX	
PT	PT	Catalytic RNA having aminocyclating activity comprises a tRNA-like and a
PT	PT	ribozyme domain and is useful to produce non-natural amino acids for
XX	XX	biomedical and therapeutic purposes.
XX	XX	
PS	PS	Claim 1; Page 34; 38pp; English.
XX	XX	
CC	CC	The sequences given in AAC85923-36 are catalytic RNA molecules,
CC	CC	designated cis-aminocyclating RNA, which comprise a tRNA-like and a
CC	CC	ribozyme domain, and which have the ability to selectively amino- acylate
CC	CC	their own 3' termini with specific amino acids. The ribozyme domain has
CC	CC	the catalytic activity and also confers amino acid specificity for
CC	CC	aminocyclation. The catalytic RNA molecules are used to aminocyclate tRNA
CC	CC	-like molecules in cis or in trans for the production of proteins
CC	CC	containing non-natural amino acids that may be of use for biomedical or
CC	CC	therapeutic purposes. The catalytic RNA molecules were constructed by
CC	CC	applying selection to a randomly synthesised RNA pool. By attaching the
CC	CC	CNA molecules of this pool to the 5' end of a tRNA-like molecule and
CC	CC	contacting the complex with a substrate molecule (natural or non-natural
CC	CC	amino acid), self- aminocyclating RNA molecules were identified
XX	XX	
SQ	SQ	Sequence 149 BP; 30 A; 38 C; 45 G; 0 T; 36 U; 0 Other;
XX	XX	
Query Match	90.9%;	Score 136.4; DB 4; Length 149;
Best Local Similarity	98.7%;	Pred. No. 8.6e-40;
Matches 148; Conservative	0;	Mismatches 1; Indels 1; Gaps 1
DB		
QY		1 GGAUCGUCAGUCAGUAGAUUUCGCGAGGCCUUCUCACUACGUGGUGUACGUGU 60
DB		1 GGAUCGUCAGUCAGUAGAUUUCGCGA-GCCUUCUCACUACGUGUAGUACGUGU 59
QY		61 UGGCGUUAUGUGCGGAGUACUACGUGUGUGUUAUCCCAAGGUAACGGACCGGACA 120
DB		60 UGGCGUUAUGUGCGGAGUACUACGUGUGUGUUAUCCCAAGGUAACGGACCGGACA 119

Oy 121 UUCGAGAUUCCGUAUCCGACGCA 150
 Db 120 UCCGAGAUUCCGUAUCCGACGCA 149

RESULT 5

AAC85930 ID AAC85930 standard; RNA; 149 BP.

XX AAC85930;

XX 22-AUG-2001 (first entry)

XX Pre-5 catalytic RNA.

XX KW Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
 amino acid; aminoacylation; ss.

XX OS Synthetic.

XX Key Location/Qualifiers

FT misc_RNA 1..88

FT /tag= a /label= 5' leader domain

FT tRNA 89..149

FT /tag= b /label= tRNA domain

FT stem_loop 98..111

FT D_loop 118..121

FT /tag= c /label= D-loop

FT /tag= d /label= V-loop

FT stem_loop 122..138

FT /tag= e /label= T-loop

XX WO200138582-A1.

XX 31-MAY-2001.

XX 22-NOV-2000; 2000WO-US032184.

XX 24-NOV-1999; 99US-0167331P.

XX 28-JUN-2000; 2000US-0214382P.

XX (UYNV) UNIV NEW YORK STATE RES FOUND.

XX Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzaghi N,

XX WPI; 2001-381295/40.

XX Claim 1; Page 34; 38pp; English.

XX The sequences given in AAC85923-36 are catalytic RNA molecules,
 CC designed cis-aminoacylating RNA, which comprise a tRNA-like and a
 CC ribozyme domain, and which have the ability to selectively amino- acylate
 CC their own 3' termini with specific amino acids. The ribozyme domain has
 CC the catalytic activity and also confers amino acid specificity for
 CC aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
 CC -like molecules in cis or in trans for the production of proteins
 CC containing non-natural amino acids that may be of use for biomedical or
 CC therapeutic purposes. The catalytic RNA molecules were constructed by
 CC applying selection to a randomly synthesised RNA pool. By attaching the
 CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
 CC connecting the complex with a substrate molecule (natural or non-natural
 CC amino acid), self- aminoacylating RNA molecules were identified

SQ Sequence 149 BP; 28 A; 39 C; 46 G; 0 T; 35 U; 1 Other;

Query Match 90.3%; Score 135.4; DB 4; Length 149;
 Best Local Similarity 98.0%; Pred. No. 2e-39; 2; Indels 1; Gaps 1;
 Matches 147; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Oy 1 GGAUCCGACGAGUAGAUUUUCCGAGGCCUUCUACUAACGUGGUGUAGUGGUAU 60

Db 1 GGAUCCGACGAGUAGAUUUUCCGAGGCCUUCUACUAACGUGGUGUAGUGGUAU 59

Oy 61 UGCGGUUAGUGGCGGAGUUCUACUACGUGUGUUAUCCCAAGGUAACGAGACGACA 120

Db 60 UGCGGUUAGUGGCGGAGUUCUACUACGUGUGUUAUCCCAAGGUAACGAGACGACA 119

Oy 121 UUCGAGAUUCCGUAUCCGACGCA 150

Db 120 UCCGAGAUUCCGUAUCCGACGCA 149

RESULT 6

AAC85929 ID AAC85929 standard; RNA; 149 BP.

XX AAC85929;

XX 22-AUG-2001 (first entry)

XX Pre-22 catalytic RNA.

XX KW Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
 amino acid; aminoacylation; ss.

XX OS Synthetic.

XX Key Location/Qualifiers

FT misc_RNA 1..88

FT /tag= a /label= 5' leader domain

FT tRNA 89..149

FT /tag= b /label= tRNA domain

FT stem_loop 98..111

FT /tag= c /label= D-loop

FT /tag= d /label= V-loop

FT stem_loop 122..138

FT /tag= e /label= T-loop

XX WO200138582-A1.

XX 31-MAY-2001.

XX 22-NOV-2000; 2000WO-US032184.

XX 24-NOV-1999; 99US-0167331P.

XX 28-JUN-2000; 2000US-0214382P.

XX (UYNV) UNIV NEW YORK STATE RES FOUND.

XX Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzaghi N,

XX WPI; 2001-381295/40.

XX Claim 1; Page 33; 38pp; English.

XX The sequences given in AAC85923-36 are catalytic RNA molecules,

CC designated cis-aminoacylating RNA, which comprise a tRNA-like and a
 CC ribozyme domain, and which have the ability to selectively amino-acylate
 CC their own 3' terminus with specific amino acids. The ribozyme domain has
 CC the catalytic activity and also confers amino acid specificity for
 CC aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
 CC -like molecules in cis or in trans for the production of proteins
 CC containing non-natural amino acids that may be of use for biomedical or
 CC therapeutic purposes. The catalytic RNA molecules were constructed by
 CC applying selection to a randomly synthesised RNA pool. By attaching the
 CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
 CC contacting the complex with a substrate molecule (natural or non-natural
 CC amino acid), self- aminoacylating RNA molecules were identified

XX Sequence 149 BP; 28 A; 37 C; 47 G; 0 T; 37 U; 0 Other;

Query Match 89.9%; Score 134.8; DB 4; Length 149;
 Best Local Similarity 98.0%; Pred. No. 3.3e-39;
 Matches 147; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 GGAUUGCAGUGCAUUGAGAUUUGCGAGCCUUCUACUAACGGUGGUGUCAUGGUAU 60
 Db 1 GGAUUGCAGUGCAUUGAGAUUUGCGAGCCUUCUACUAACGGUGGUGUCAUGGUAU 59
 Qy 61 UGGCGUUAAGUGCGGGAGUCUAACGUGUGUAUUGCCCAAGGUAACGGACCGGACA 120
 Db 60 UGGCGUUAAGUGCGGGAGUCUAACGUGUGUAUUGCCCAAGGUAACGGACCGGACA 119
 Qy 121 UUCGAGAUUCGAUUCUCCUGACCGACGCA 150
 Db 120 UUCGAGAUUCGAUUCUCCUGACCGACGCA 149

RESULT 7

AAc85926 standard; RNA; 150 BP.

AC AAc85926;

DT 22-AUG-2001 (first entry)

XX Pre-36 catalytic RNA.

DE Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
 KW amino acid; aminoacylation; ss.

XX Synthetic.

OS Key Location/Qualifiers

FT misc_RNA

FT 1..88
 FT /tag= a
 FT /label= 5' leader domain

FT tRNA

FT 89..150
 FT /tag= b
 FT /label= tRNA domain

FT stem_loop

FT 98..112
 FT /tag= C
 FT /label= D-loop

FT D_loop

FT 120..123
 FT /tag= d
 FT /label= V-loop

FT stem_loop

FT 124..140
 FT /tag= e
 FT /label= T-loop

XX WO200138582-A1.

XX 31-MAY-2001.

PD 22-NOV-2000; 2000WO-US032184.

PR 24-NOV-1999; 99US-0167331P.
 PR 28-JUN-2000; 2000US-0214382P.

PA (UWNY) UNIV NEW YORK STATE RES FOUND.

XX PA Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzagni N;

XX WPI; 2001-381295/40.

PT Catalytic RNA having aminoacylating activity comprises a tRNA-like and a
 PT ribozyme domain and is useful to produce non-natural amino acids for
 PT biomedical and therapeutic purposes.

XX Claim 1; Page 33; 38pp; English.

CC The sequences given in AAC85923-36 are catalytic RNA molecules,
 CC designated cis-aminoacylating RNA, which comprise a tRNA-like and a
 CC ribozyme domain, and which have the ability to selectively amino-acylate
 CC their own 3' terminus with specific amino acids. The ribozyme domain has
 CC the catalytic activity and also confers amino acid specificity for
 CC aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
 CC -like molecules in cis or in trans for the production of proteins
 CC containing non-natural amino acids that may be of use for biomedical or
 CC therapeutic purposes. The catalytic RNA molecules were constructed by
 CC applying selection to a randomly synthesised RNA pool. By attaching the
 CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
 CC contacting the complex with a substrate molecule (natural or non-natural
 CC amino acid), self- aminoacylating RNA molecules were identified

XX Sequence 150 BP; 29 A; 38 C; 46 G; 0 T; 37 U; 0 Other;

Query Match 84.7%; Score 127; DB 4; Length 150;
 Best Local Similarity 98.7%; Pred. No. 2.4e-36;
 Matches 149; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 GGAUUGCAGUGCAUUGAGAUUUGCGAGCCUUCUACUAACGGUGGUGUCAUGGUAU 60
 Db 1 GGAUUGCAGUGCAUUGAGAUUUGCGAGCCUUCUACUAACGGUGGUGUCAUGGUAU 59
 Qy 61 UGGCGUUAAGUGCGGGAGUCUAACGUGUGUAUUGCCCAAGGUAACGGACCGGACA-C 119
 Db 60 UGGCGUUAAGUGCGGGAGUCUAACGUGUGUAUUGCCCAAGGUAACGGACCGGAUC 119
 Qy 120 AUUUGAGAUUCGAUUCUCCUGACCGACGCA 150
 Db 120 AUUUGAGAUUCGAUUCUCCUGACCGACGCA 150

RESULT 8

AAc85933 standard; RNA; 148 BP.

AC AAc85933;

DT 22-AUG-2001 (first entry)

XX Pre-23 catalytic RNA.

DE Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
 KW amino acid; aminoacylation; ss.

XX Synthetic.

OS Key Location/Qualifiers

FT misc_RNA

FT 1..88
 FT /tag= a
 FT /label= 5' leader domain

FT tRNA

FT 89..148
 FT /tag= b
 FT /label= tRNA domain

FT stem_loop

FT 98..110
 FT /tag= C
 FT /label= D-loop

FT D_loop

FT 117..120
 FT /tag= d
 FT /label= V-loop

```

FT      stem_loop      121..137
FT      /*tag= e
FT      /label= T-loop
XX
XX      WO200138582-A1.
XX
XX      31-MAY-2001.
XX
XX      22-NOV-2000; 2000WO-US032184.
XX
XX      24-NOV-1999; 99US-0167331P.
XX      28-JUN-2000; 2000US-0214382P.
XX
XX      (UYNV ) UNIV NEW YORK STATE RES FOUND.
XX
XX      Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzagnt N;
XX      WPI; 2001-381295/40.
XX
XX      Catalytic RNA having aminoacylating activity comprises a tRNA-like and a
XX      ribozyme domain and is useful to produce non-natural amino acids for
XX      biomedical and therapeutic purposes.
XX
XX      Claim 1; Page 34; 38pp; English.
XX
XX      The sequences given in AAC85923-36 are catalytic RNA molecules,
XX      designated cis-aminoacylating RNA, which comprise a tRNA-like and a
XX      ribozyme domain, and which have the ability to selectively amino- acylate
XX      their own 3' terminus with specific amino acids. The ribozyme domain has
XX      the catalytic activity and also confers amino acid specificity for
XX      aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
XX      -like molecules in cis or in trans for the production of proteins
XX      containing non-natural amino acids that may be of use for biomedical or
XX      therapeutic purposes. The catalytic RNA molecules were constructed by
XX      applying selection to a randomly synthesised RNA pool. By attaching the
XX      RNA molecules of this pool to the 5' end of a tRNA-like molecule and
XX      contacting the complex with a substrate molecule (natural or non-natural
XX      amino acid), self- aminoacylating RNA molecules were identified
XX
XX      Sequence 148 BP; 29 A; 37 C; 46 G; 0 T; 36 U; 0 Other;
XX
XX      Query Match      84.0%; Score 126; DB 4; Length 148;
XX      Best Local Similarity 98.7%; Pred. No. 5.5e-36;
XX      Matches 148; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
XX
XX      QY      1 GGAUCGUAUGCAUUGAGAUUUCGAGGCCCTUCUACUAACGGUGGUAUGGGU 60
XX      DB      1 GGAUCGUAUGCAUUGAGAUUUCGCA-GCCCTUCUACUAACGGUGGUAUGGGU 59
XX
XX      QY      61 UGGCGUAGUGGCGGAGUAGCUAUCGUGUGUUAUCCCAAGGUAAGGACCGGACA 120
XX      DB      60 UGGCGUAGUGGCGGAGUAGCUAUCGUGUGUUAU-CCCAAGGUAAGGACCGGACA 118
XX
XX      QY      121 UUCGAGAUUUCGAAUUCUUGUAACCGGACCA 150
XX      DB      119 UUCGAGAUUUCGAAUUCUUGUAACCGGACCA 148
XX
XX      RESULT 9
XX      AAC85932
XX      ID      AAC85932 standard; RNA; 150 BP.
XX
XX      AC      AAC85932;
XX
XX      XX      22-AUG-2001 (first entry)
XX
XX      DE      Pre-8 catalytic RNA.
XX
XX      KM      Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
XX      KW      amino acid; aminoacylation; ss.
XX      OS      Synthetic.
XX

```

```

FH      Key      Location/Qualifiers
FH      mlec_RNA      1..89
FH      /*tag= a
FH      /label= 5' leader domain
FH      tRNA      90..150
FH      /*tag= b
FH      /label= tRNA domain
FH      stem_loop      99..112
FH      /*tag= c
FH      /label= D-loop
FH      D_loop      119..122
FH      /*tag= d
FH      /label= V-loop
FH      stem_loop      123..139
FH      /*tag= e
FH      /label= T-loop
XX
XX      WO200138582-A1.
XX
XX      31-MAY-2001.
XX
XX      22-NOV-2000; 2000WO-US032184.
XX
XX      24-NOV-1999; 99US-0167331P.
XX      28-JUN-2000; 2000US-0214382P.
XX
XX      (UYNV ) UNIV NEW YORK STATE RES FOUND.
XX
XX      Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzagnt N;
XX      WPI; 2001-381295/40.
XX
XX      Catalytic RNA having aminoacylating activity comprises a tRNA-like and a
XX      ribozyme domain and is useful to produce non-natural amino acids for
XX      biomedical and therapeutic purposes.
XX
XX      Claim 1; Page 34; 38pp; English.
XX
XX      The sequences given in AAC85923-36 are catalytic RNA molecules,
XX      designated cis-aminoacylating RNA, which comprise a tRNA-like and a
XX      ribozyme domain, and which have the ability to selectively amino- acylate
XX      their own 3' terminus with specific amino acids. The ribozyme domain has
XX      the catalytic activity and also confers amino acid specificity for
XX      aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
XX      -like molecules in cis or in trans for the production of proteins
XX      containing non-natural amino acids that may be of use for biomedical or
XX      therapeutic purposes. The catalytic RNA molecules were constructed by
XX      applying selection to a randomly synthesised RNA pool. By attaching the
XX      RNA molecules of this pool to the 5' end of a tRNA-like molecule and
XX      contacting the complex with a substrate molecule (natural or non-natural
XX      amino acid), self- aminoacylating RNA molecules were identified
XX
XX      Sequence 150 BP; 29 A; 37 C; 45 G; 0 T; 39 U; 0 Other;
XX
XX      Query Match      82.9%; Score 124.4; DB 4; Length 150;
XX      Best Local Similarity 89.3%; Pred. No. 2.1e-35;
XX      Matches 134; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
XX
XX      QY      1 GGAUCGUAUGCAUUGAGAUUUCGAGGCCCTUCUACUAACGGUGGUAUGGGU 60
XX      DB      1 GGAUCGUAUGCAUUGAGAUUUCGAGGCCCTUCUACUAACGGUGGUAUGGGU 60
XX
XX      QY      61 UGGCGUAGUGGCGGAGUAGCUAUCGUGUGUUAUCCCAAGGUAAGGACCGGACA 120
XX      DB      61 UGGCGUAGUGGCGGAGUAGCUAUCGUGUGUUAUCCCAAGGUAAGGACCGGACA 120
XX
XX      QY      121 UUCGAGAUUUCGAAUUCUUGUAACCGGACCA 150
XX      DB      121 UUCGAGAUUUCGAAUUCUUGUAACCGGACCA 150
XX
XX      RESULT 10
XX      AAC85923

```

```

ID AAC85923 standard; RNA; 146 BP.
XX
XX AAC85923;
AC
XX
XX 22-AUG-2001 (first entry)
DT
XX
XX Pre-12 catalytic RNA.
DE
XX
XX Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
KM amino acid; aminoacylation; ss.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH 1..85
FT misc_RNA /*tag= a
FT /*label= 5', leader domain
FT 86..146
FT tRNA /*tag= b
FT /*label= tRNA domain
FT 95..108
FT stem_loop /*tag= c
FT /*label= D-loop
FT 115..118
FT D_loop /*tag= d
FT /*label= V-loop
FT 119..135
FT stem_loop /*tag= e
FT /*label= T-loop
FT
FT
FT
FT
XX
XX MO200138582-A1.
XX
XX 31-MAY-2001.
PD
XX
XX 22-NOV-2000; 2000MO-US032184.
PF
XX
XX 24-NOV-1999; 99US-0167331P.
PR
XX 28-JUN-2000; 2000US-0214382P.
XX
XX (UYNV ) UNIV NEW YORK STATE RES FOUND.
XX
XX Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzaghi N;
XX
XX WPI; 2001-381295/40.
XX
XX
XX Catalytic RNA having aminoacylating activity comprises a tRNA-like and a
XX ribozyme domain and is useful to produce non-natural amino acids for
XX biomedical and therapeutic purposes.
XX
XX Claim 1; Page 32; 38pp; English.
XX
XX The sequences given in AAC85923-36 are catalytic RNA molecules,
XX designated cis-aminoacylating RNA, which comprise a tRNA-like and a
XX ribozyme domain, and which have the ability to selectively amino-acylate
XX their own 3' terminus with specific amino acids. The ribozyme domain has
XX the catalytic activity and also confers amino acid specificity for
XX aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
XX -like molecules in cis or in trans for the production of proteins
XX containing non-natural amino acids that may be of use for biomedical or
XX therapeutic purposes. The catalytic RNA molecules were constructed by
XX applying selection to a randomly synthesised RNA pool. By attaching the
XX RNA molecules of this pool to the 5' end of a tRNA-like molecule and
XX contacting the complex with a substrate molecule (natural or non-natural
XX amino acid), self- aminoacylating RNA molecules were identified
XX
XX Sequence 146 BP; 28 A; 37 C; 46 G; 0 T; 35 U; 0 Other;
SQ
XX
XX Query Match 81.3%; Score 122; DB 4; Length 146;
XX Best Local Similarity 97.3%; Pred. No. 1.6e-34;
XX Matches 146; Conservative 0; Mismatches 0; Indels 4; Gaps 2;
XX
XX 1 GGAGUCGACGUGCAUUGAGAUUUCGAGCCCUUCUCACUACGUGGUGUACUGGUU 60
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

```

DB 1 GGAGUCGACGUGCAUUGAGAUUUCGCA-GCCUUCUCACUACGUGGUGUACUGGUU 59
XX
XX QY 61 UGCGCGUUGGUGCGGGAGUGCAUACGUGGUGGUUACCCAGGUAACGGACA 120
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 60 UGCGCGUUGAGUGCGGGAGUG---CUACGUGGUGUACCCAGGUAACGGACA 116
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 UUGGAGAUUUGCAUUCGUGUACCGGAGCCA 150
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 117 UUGGAGAUUUGCAUUCGUGUACCGGAGCCA 146
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX
XX RESULT 11
XX AAC85924
XX ID AAC85924 standard; RNA; 151 BP.
XX
XX AAC85924;
AC
XX
XX 22-AUG-2001 (first entry)
DT
XX
XX Pre-38 catalytic RNA.
DE
XX
XX Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
KM amino acid; aminoacylation; ss.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH 1..89
FT misc_RNA /*tag= a
FT /*label= 5', leader domain
FT 90..151
FT tRNA /*tag= b
FT /*label= tRNA domain
FT 99..113
FT stem_loop /*tag= c
FT /*label= D-loop
FT 121..124
FT D_loop /*tag= d
FT /*label= V-loop
FT 125..141
FT stem_loop /*tag= e
FT /*label= T-loop
FT
FT
FT
FT
XX
XX MO200138582-A1.
XX
XX 31-MAY-2001.
PD
XX
XX 22-NOV-2000; 2000MO-US032184.
PF
XX
XX 24-NOV-1999; 99US-0167331P.
PR
XX 28-JUN-2000; 2000US-0214382P.
XX
XX (UYNV ) UNIV NEW YORK STATE RES FOUND.
XX
XX Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzaghi N;
XX
XX WPI; 2001-381295/40.
XX
XX Catalytic RNA having aminoacylating activity comprises a tRNA-like and a
XX ribozyme domain and is useful to produce non-natural amino acids for
XX biomedical and therapeutic purposes.
XX
XX Claim 1; Page 32; 38pp; English.
XX
XX The sequences given in AAC85923-36 are catalytic RNA molecules,
XX designated cis-aminoacylating RNA, which comprise a tRNA-like and a
XX ribozyme domain, and which have the ability to selectively amino-acylate
XX their own 3' terminus with specific amino acids. The ribozyme domain has
XX the catalytic activity and also confers amino acid specificity for
XX aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
XX -like molecules in cis or in trans for the production of proteins
XX containing non-natural amino acids that may be of use for biomedical or
XX therapeutic purposes. The catalytic RNA molecules were constructed by
XX
XX

```

CC applying selection to a randomly synthesised RNA pool. By attaching the
 CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
 CC contacting the complex with a substrate molecule (natural or non-natural
 CC amino acid), self- aminoacylating RNA molecules were identified

XX Sequence 151 BP; 29 A; 38 C; 45 G; 0 T; 38 U; 1 Other;

Query Match 77.1%; Score 115.6; DB 4; Length 151;
 Best Local Similarity 89.4%; Pred. No. 3.5e-32;
 Matches 135; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

Qy 1 GGAUCGUCAGUCAGUAGAUUUCGAGGCCCUUCUCACUAACGGUGGUCAGUGGUAU 60
 Db 1 GGAUCGUCAGUACAUUGAGAUUUCGAGCCCUUCUCACUAACGGUGGUCAGUGGUAU 60
 Qy 61 UGGGUGUAGUGUGCGGAGUACUAACGUGUGUAGUCCCAAGGUAACGGACCGGA-C 119
 Db 61 UGGGUGUAGUGUGCGGAGUACUAACGUGUGUAGUCCCAAGGUAACGGACCGGAUUC 120
 Qy 120 AUGCAGAUUCGAAUCCUCGUAACCGCAGCCA 150
 Db 121 AUGCAGAUUCGAAUCCUCGUAACCGCAGCCA 151

RESULT 12

AAC85937
 ID AAC85937 standard; RNA; 160 BP.

XX AAC85937;

XX 22-AUG-2001 (first entry)

XX H2 Leu catalytic RNA.

XX Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
 KW amino acid; aminoacylation; ss.

XX Synthetic.

XX WO200138582-A1.

XX 31-MAY-2001.

XX 22-NOV-2000; 2000WO-US032184.

XX 24-NOV-1999; 99US-0167331P.

XX 28-JUN-2000; 2000US-0214382P.

XX (UTNY) UNIV NEW YORK STATE RES FOUND.

XX Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzaghi N;

XX WPI; 2001-381295/40.

XX Catalytic RNA having aminoacylating activity comprises a tRNA-like and a
 PT ribozyme domain and is useful to produce non-natural amino acids for
 PT biomedical and therapeutic purposes.

XX Example 5; Page 35; 38pp; English.

XX The sequences given in AAC85937-38 are catalytic RNA molecules,
 CC designated cis-aminoacylating RNA, which comprise a tRNA-like and a
 CC ribozyme domain, and which have the ability to selectively amino-acylate
 CC their own 3' termini with specific amino acids. These catalytic RNA's
 CC are leucine specific cis aminoacylating RNA's. The ribozyme domain has
 CC the catalytic activity and also confers amino acid specificity for
 CC aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
 CC -like molecules in cis or in trans for the production of proteins
 CC containing non-natural amino acids that may be of use for biomedical or
 CC therapeutic purposes. The catalytic RNA molecules were constructed by
 CC applying selection to a randomly synthesised RNA pool. By attaching the
 CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
 CC contacting the complex with a substrate molecule (natural or non-natural

CC amino acid), self- aminoacylating RNA molecules were identified

XX Sequence 160 BP; 34 A; 37 C; 49 G; 0 T; 40 U; 0 Other;

Query Match 47.1%; Score 70.6; DB 4; Length 160;
 Best Local Similarity 74.5%; Pred. No. 1.1e-15;
 Matches 120; Conservative 0; Mismatches 29; Indels 12; Gaps 2;

Qy 1 GGAUCGUCAGUCAGUAGAUUUCGAGGCCCUUCUCACUAACGGUGGUCAGUGGUAU 60
 Db 1 GGAUCGUCAGUCAGUAGAGA-UGCCCAAAGCCCUUCUCACUUCGUGGUCAGUGGUAU 59
 Qy 61 UGGGUGUAGUGUGCGGAGUACUAACGUGUGUAGUCCCAAGGUAACGG- 111
 Db 60 UUGGUGUAGUGAGGAUUGCUAGUAGCGGUGGUAUCCAAAGGUGUAAGGACCGGAUUC 119
 Qy 112 --GACCGACAUUCGAAUUCGAAUUCGUAACCGCAGCCA 150
 Db 120 UAAUCCGACAUUCGAGGUGGAAUCCUCGUAACCGCAGCCA 160

RESULT 13

AAC85934
 ID AAC85934 standard; RNA; 73 BP.

XX AAC85934;

XX 22-AUG-2001 (first entry)

XX V1 variant of ocrRNA.

XX Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
 KW amino acid; aminoacylation; ss.

XX Synthetic.

XX Key Location/Qualifiers

XX stem_loop 10..23

XX stem_loop 25..41

XX D_loop 42..46

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX stem_loop 47..62

XX The sequences given in AAC85923-36 are catalytic RNA molecules,
 CC designated cis-aminoacylating RNA, which comprise a tRNA-like and a

CC ribozyme domain, and which have the ability to selectively amino- acylate
CC their own 3' terminus with specific amino acids. The ribozyme domain has
CC the catalytic activity and also confers amino acid specificity for
CC aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
CC -like molecules in cis or in trans for the production of proteins
CC containing non-natural amino acids that may be of use for biomedical or
CC therapeutic purposes. The catalytic RNA molecules were constructed by
CC applying selection to a randomly synthesised RNA pool. By attaching the
CC RNA molecules of this pool to the 5' end of a tRNA-like molecule and
CC contacting the complex with a substrate molecule (natural or non-natural
CC amino acid), self- aminoacylating RNA molecules were identified

Sequence 73 BP; 17 A; 21 C; 20 G; 0 T; 15 U; 0 Other;

Query Match 25.0%; Score 39; DB 4; Length 73;
Best Local Similarity 83.6%; Pred. No. 0.00031;
Matches 61; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 90 GGUGUAUCCCCCAAGGUAACGGAGCCGGA-----CAUUCGAGAUUCGAAUCCU 137
DB 1 GGUGUAUCCCCCAAGGUAACGGAGCCGGAUUCUAAUCCGCAUUCGAGAUUCGAAUCCU 60
QY 138 CGUACCGCAGCCA 150
DB 61 CGUACCGCAGCCA 73

RESULT 14

ADCI6643 standard; tRNA; 74 BP.

ADCI6643;

18-DEC-2003 (first entry)

Aminoacylation RNA molecule related vi-tRNA.

ribozyme; aminoacylate; tRNA; non-cognate; catalytic RNA molecule; cis;
aminoacylation; trans; proteomic; ss.

Unidentified.

WO2003070740-A1.

28-AUG-2003.

18-FEB-2003; 2003WO-US005007.

15-FEB-2002; 2002US-0357424P.

(UNYV) UNIV NEW YORK STATE RES FOUND.

Suga H, Murakami H, Saito H;

WPI; 2003-748198/70.

New polynucleotide, useful for preparing peptides containing non-cognate
amino acids, encodes ribozyme that can aminoacylate tRNA with such amino
acids.

Disclosure; SEQ ID NO 3; 85pp; English.

The invention relates to a novel polynucleotide comprising a sequence
encoding a ribozyme that can aminoacylate tRNA with a non-cognate amino
acid. Ribozymes encoded by the polynucleotide of the invention are used
to prepare polypeptides that contain non-cognate, including non-natural,
amino acids. The invention more specifically provides catalytic RNA
molecules having cis aminoacylation activity with a catalytic and
aminoacylation domain, or an RNA molecule with trans aminoacylation
activity with only a catalytic domain. The products of the invention are
potentially useful for biomedical and therapeutic use, e.g. for probing
the structure and function of proteins; preparation of peptide libraries
and in proteomics. This polynucleotide sequence represents a vi-tRNA

CC structure relating to the RNA molecule with aminoacylation activity of
CC the invention.

Sequence 74 BP; 16 A; 23 C; 21 G; 0 T; 14 U; 0 Other;

Query Match 25.3%; Score 38; DB 9; Length 74;
Best Local Similarity 82.4%; Pred. No. 0.00072;
Matches 61; Conservative 0; Mismatches 0; Indels 13; Gaps 1;

QY 90 GGUGUAUCCCCCAAGGUAACGGAGCCGGA-----CAUUCGAGAUUCGAAUCC 136
DB 1 GGUGUAUCCCCCAAGGUAACGGAGCCGGAUUCUCCGCAUUCGAGAUUCGAAUCC 60
QY 137 UCGUACCGCAGCCA 150
DB 61 UCGUACCGCAGCCA 74

RESULT 15

AAC85938 standard; RNA; 156 BP.

AAC85938;

22-AUG-2001 (first entry)

D1-Leu catalytic RNA.

Catalytic RNA; cis-aminoacylating RNA; tRNA-like domain; ribozyme domain;
amino acid; aminoacylation; ss.

Synthetic.

WO200138582-A1.

31-MAY-2001.

22-NOV-2000; 2000WO-US032184.

24-NOV-1999; 99US-0167331P.

28-JUN-2000; 2000US-0214382P.

(UNYV) UNIV NEW YORK STATE RES FOUND.

Hiroaki S, Dimitrios K, Hirohide S, Lee N, Bonzagani N;

WPI; 2001-381295/40.

Catalytic RNA having aminoacylating activity comprises a tRNA-like and a
ribozyme domain and is useful to produce non-natural amino acids for
biomedical and therapeutic purposes.

Example 5; Page 35; 38pp; English.

The sequences given in AAC85937-38 are catalytic RNA molecules,
designated cis-aminoacylating RNA, which comprise a tRNA-like and a
ribozyme domain, and which have the ability to selectively amino- acylate
their own 3' terminus with specific amino acids. These catalytic RNA's
are leucine specific cis aminoacylating RNA's. The ribozyme domain has
the catalytic activity and also confers amino acid specificity for
aminoacylation. The catalytic RNA molecules are used to aminoacylate tRNA
-like molecules in cis or in trans for the production of proteins
containing non-natural amino acids that may be of use for biomedical or
therapeutic purposes. The catalytic RNA molecules were constructed by
applying selection to a randomly synthesised RNA pool. By attaching the
RNA molecules of this pool to the 5' end of a tRNA-like molecule and
contacting the complex with a substrate molecule (natural or non-natural
amino acid), self- aminoacylating RNA molecules were identified

Sequence 156 BP; 31 A; 27 C; 55 G; 0 T; 43 U; 0 Other;

Query Match 24.3%; Score 36.4; DB 4; Length 156;
Best Local Similarity 57.1%; Pred. No. 0.0034;

	Matches	89;	Conservative	0;	Mismatches	61;	Indels	6;	Gaps	1;
Qy	1	GGATCGGACAGGCAUUGAGAUUUCCGACAGCCCUUCACUACGAGGAGUATGGGUU	60							
Db	1	GGATCGGACAGGCAUUGAGAUUUCCGACAGCCCUUCACUACGAGGAGUATGGGUU	60							
Qy	61	UGGCGUUAGGUGCGGAGUGCUACUACGCGUGGUGGUUCCCCAGGGUACGG-----GAC	114							
Db	61	UGGUTUGGGGGGUUAVGCGUGAGUUCUUGGGUGGUUCCAGGGGUUAGGGAUUCUAAAU	120							
Qy	115	CGGACAUUCGAGAUUCCGAAUCCUCGUACCGCAGCCA	150							
Db	121	CGACAUUCCGAGGUCCGAAUCCUCGUACCGCAGCCA	156							

Search completed: October 8, 2004, 02:53:08
 Job time : 386.333 secs

Qy 30 GCCCTTCACUAA CGUGGAGUAVUGGCGUAGCGGGAGUCUACUAGCU 89
||| | : | ||| : | : ||| ||| : | ||
Db 206 GCCCCCCCGGAGCGGGGATCCGGGTCGCTCTAGTGTGATGATGCTCCCACTGC 147

Oy 90 GGUGUAUCCCCAGAGGAGCGAGACCGGACAUUCGAGAUUCUCUGACCGGACCC 149
Db 146 TTCGCTCCAGAAAGTGTCCGCTCAGCCCGGTTGAGACTGAGTCCGCTAGCCGCTGCC 87

RESULT 2

US-08-242-098-28
Sequence 28, Application US/08242098
Patent No. 5691185
GENERAL INFORMATION:
APPLICANT: DICKELLY, Franciscas
APPLICANT: JOHANSEN, Eric
APPLICANT: NILSSON, Dan
APPLICANT: HANSEN, Egon
APPLICANT: STROMAN, Per
TITLE OF INVENTION: Lactic Acid Bacterial Suppressor Mutants
TITLE OF INVENTION: and Their Use as Selective Markers and as Means of
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,098
FILING DATE: 13-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,390
FILING DATE: 08-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30307/141/PLVI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-242-098-28

Query Match 18.3%; Score 27.4; DB 1; Length 75;
Best Local Similarity 67.6%; Pred. No. 0.38;
Matches 25; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Oy 114 CCGGACAUUCGAGAUUCGAGAUUCUCUGACCGGACCA 150
Db 39 CCGGATTCGAGGTTGGAATCTCGTACCCGACCA 75

RESULT 3

US-09-489-039A-6219/c
Sequence 6219, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A

CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 6219
LENGTH: 183
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6219

Query Match 18.3%; Score 27.4; DB 4; Length 183;
Best Local Similarity 67.6%; Pred. No. 0.54;
Matches 25; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Oy 114 CCGGACAUUCGAGAUUCGAGAUUCUCUGACCGGACCA 150
Db 40 CCGGATTCGAGGTTGGAATCTCGTACCCGACCA 4

RESULT 4

US-09-489-039A-5600/c
Sequence 5600, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 5600
LENGTH: 186
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5600

Query Match 18.3%; Score 27.4; DB 4; Length 186;
Best Local Similarity 67.6%; Pred. No. 0.54;
Matches 25; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Oy 114 CCGGACAUUCGAGAUUCGAGAUUCUCUGACCGGACCA 150
Db 130 CCGGATTCGAGGTTGGAATCTCGTACCCGACCA 94

RESULT 5

US-09-489-039A-5979/c
Sequence 5979, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 5979
LENGTH: 360
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5979

Query Match 18.3%; Score 27.4; DB 4; Length 360;
Best Local Similarity 67.6%; Pred. No. 0.77;
Matches 25; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Oy 114 CCGGACAUUCGAGAUUCGAGAUUCUCUGACCGGACCA 150

Db 245 CCGGCATTCCGAGGTTGGAATCTCTGTAACCCAGCCA 209

```

RESULT 6
US-09-790-988-1/c
: Sequence 1, Application US/09790988
: Patent No. 6632935
: GENERAL INFORMATION:
: APPLICANT: SHIGENOBU, SHUJI
: APPLICANT: MATANABE, HIDEMI
: APPLICANT: HATTORI, MASAHIRA
: APPLICANT: SAKAKI, YOSHIYUKI
: TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
: FILE REFERENCE: 081356/0159
: CURRENT APPLICATION NUMBER: US/09/790,988
: CURRENT FILING DATE: 2001-02-23
: PRIOR APPLICATION NUMBER: JP2000-107160
: PRIOR FILING DATE: 2000-04-07
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 1
: LENGTH: 640681
: TYPE: DNA
: ORGANISM: Buchnera sp.
US-09-790-988-1

```

```

RESULT 7
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; US-09-103-840A-2

```

```

RESULT 8
US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

RESULT 9
US-09-252-991A-14986/C
: Sequence 14986, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 14986
: LENGTH: 444
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-14986

```

Dy 92 UGUUAUCCCCAGGGUACGCGACCGGACAUUUCGAUAUUCGAUCCUGUACCGCAGC 148
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
3343043 TGGCAACACGACTATTTCTGGTTCAGGCATTCTAGGTTTCGAGTCTCGGTATCCCAGC 3342987

RESULT 10
US-09-252-991A-14858/c
; Sequence 14858, Application US/09252991A
; Patent No. 6551795

```

Query March 17.24; Score 25.8; DB 3; Length 127;
Best Local Similarity 48.44; Prd. No. 1.9;
Matches 45; Conservative 6; Mismatches 42; Indels 0; Gaps 0;

QY      24  CCGCAGGCCCCUCACUAACGGUGGAGUCAGGUAUUGACCGUAGUGCGGAGUCCAC 83
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      34  CCGGTGGCGGTGGGCGGCAACGGGCGGCGCGGGGTGGCGGCAACGGGTGGTCTCT 93
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      84  UACGCGUGUGUAVCCCAAGGUAUGGACCG 116
      : ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      94  TCGGCAACGGCGGTGCCGCGGGGACGCGGCGC 126

RESULT 13
US-08-818-111-48
; Sequence 48, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neco, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

```

RESULT 14
US-09-056-556-48
Sequence 48, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 127 base pairs

```

1      TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
2      TUBERCULOSIS
3      NUMBER OF SEQUENCES: 350
4      CORRESPONDENCE ADDRESS:
5      ADDRESSEE: SEED AND BERRY LLP
6      STREET: 6300 Columbia Center, 701 Fifth Avenue
7      CITY: Seattle
8      STATE: Washington
9      COUNTRY: USA
10     ZIP: 98104-7092
11     COMPUTER READABLE FORM:
12     MEDIUM TYPE: floppy disk
13     COMPUTER: IBM PC compatible
14     OPERATING SYSTEM: PC-DOS/MS-DOS
15     SOFTWARE: Patent In Release #1.0, Version #1.30
16     CURRENT APPLICATION DATA:
17     APPLICATION NUMBER: US/09/072,596
18     FILING DATE: 05-MAY-1998
19     CLASSIFICATION:
20     ATTORNEY/AGENT INFORMATION:
21     NAME: Maki, David J.
22     REGISTRATION NUMBER: 31,392
23     REFERENCE/DOCKET NUMBER: 210121.417C9
24     TELECOMMUNICATION INFORMATION:
25     TELEPHONE: (206) 622-4900
26     TELEFAX: (206) 682-6031
27     INFORMATION FOR SEQ. ID NO. 48:
28     SEQUENCE CHARACTERISTICS:
29     LENGTH: 127 base pairs
30     TYPE: nucleic acid
31     STRANDEDNESS: single
32     TOPOLOGY: linear
33     IS-09-072-596-48

```

Query Match 17.2%; Score 25.8; DB 4; Length 127;
 Best Local Similarity 48.4%; Pred. No. 1.9;
 Matches 45; Conservative 6; Mismatches 42; Indels 0; Gaps 0

24 CCGCAGGCCCTTCCATCAACGAGCGGCGCAAGGAGGATGCGCGGAGCGCGGAGGCAAC 83
 34 CCGCTGCGCGCTGCGCGGACGCGCGGGCGGGGCTGCGCGCGGAGCAACCGTGGATGCTCT 93

Fr1 Oct 8 10:37:28 2004

us-09-721-414c-9.rn1

Page 6

```

Oy      84 UACGCTUGGUGUAVCCCAAGGGUACGGGACCG 116
      : | | | | | | | | | | | | | |
Db      94 TCGGCAACGGCGGTGCGCGCGGGCAAGGGCCG 126

```

Search completed: October 8, 2004, 04:46:55
Job time : 92.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: October 8, 2004, 02:42:05 ; Search time 394.667 Seconds
(without alignments)
1926.778 Million cell updates/sec

Title: US-09-721-414C-9

Perfect score: 150

Sequence: 1 ggaucgucgagcgaugaga.....gaucucgucgagcgaacca 150

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3340651 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	25.3	74	US-10-369-036B-3	Sequence 3, Appli
2	35.4	23.6	56	US-10-369-036B-12	Sequence 12, Appli
3	33	22.0	101	US-10-369-036B-8	Sequence 8, Appli
4	29.8	19.9	77	US-10-369-036B-46	Sequence 46, Appli
5	29.4	19.6	57	US-10-369-036B-1	Sequence 1, Appli
6	29.2	19.5	76	US-10-369-036B-2	Sequence 2, Appli
7	29	19.3	1143	US-10-027-632-119269	Sequence 119269,
8	29	19.3	1143	US-10-027-632-119270	Sequence 119270,
9	29	19.3	1143	US-10-027-632-119271	Sequence 119271,
10	29	19.3	1143	US-10-027-632-119269	Sequence 119269,
11	29	19.3	1143	US-10-027-632-119270	Sequence 119270,
12	29	19.3	1143	US-10-027-632-119271	Sequence 119271,
13	28.8	19.2	2050	US-10-037-270-761	Sequence 761, App
14	28.8	19.2	2050	US-10-117-722-761	Sequence 761, App

Result	Score	Query Match	Length	ID	Description
15	28.6	19.1	35	US-10-369-036B-49	Sequence 49, Appli
16	28.6	19.1	1078	US-10-128-714-1349	Sequence 1349, Ap
17	28.6	19.1	1078	US-10-128-714-6349	Sequence 6349, Ap
18	28.6	19.1	3077	US-10-128-714-349	Sequence 349, App
19	28.6	19.1	3078	US-10-128-714-5349	Sequence 5349, Ap
20	27.6	18.4	106373	US-10-388-838-96	Sequence 96, Appli
21	27.4	18.3	75	US-10-347-484C-4	Sequence 4, Appli
22	27.4	18.3	644	US-10-435-165-16	Sequence 16, Appli
23	27.4	18.3	1396	US-10-435-114-2163	Sequence 2163, Ap
24	27.2	18.1	851	US-10-767-701-9840	Sequence 9840, Ap
25	27.2	18.1	4682	US-09-764-864-1764	Sequence 1764, Ap
26	26.8	17.9	71057	US-10-087-192-409	Sequence 409, App
27	26.8	17.9	640681	US-09-799-988-1	Sequence 1, Appli
28	26.6	17.7	330	US-09-738-636-1433	Sequence 1433, Ap
29	26.6	17.7	1929	US-10-767-701-14104	Sequence 14104, A
30	26.6	17.7	6536	US-10-311-455-119	Sequence 119, App
31	26.6	17.7	6536	US-10-240-452-1	Sequence 1, Appli
32	26.6	17.7	39405	US-10-087-192-1285	Sequence 1285, Ap
33	26.6	17.7	50000	US-10-364-505-8	Sequence 8, Appli
34	26.6	17.7	50000	US-10-681-199-8	Sequence 8, Appli
35	26.6	17.7	98865	US-09-770-689A-3	Sequence 3, Appli
36	26.6	17.7	3309400	US-09-738-626-1	Sequence 1, Appli
37	26.4	17.6	480	US-09-783-590-617	Sequence 617, App
38	26.4	17.6	512	US-10-424-599-126409	Sequence 126409,
39	26.4	17.6	711	US-09-882-327-159	Sequence 159, App
40	26.4	17.6	1056	US-10-156-761-2904	Sequence 2904, Ap
41	26.4	17.6	1740	US-09-917-800A-1457	Sequence 1457, Ap
42	26.4	17.6	125746	US-10-156-761-15102	Sequence 15102, A
43	26.4	17.6	9025608	US-10-156-761-1	Sequence 1, Appli
44	26.2	17.5	100944	US-10-332-696-4	Sequence 4, Appli
45	26	17.3	572	US-09-864-761-12724	Sequence 12724, A

ALIGNMENTS

RESULT 1
US-10-369-036B-3
Sequence 3, Application US/10369036B
Publication No. US20030228593A1
GENERAL INFORMATION:
APPLICANT: Suga, Hiroaki et al.
TITLE OF INVENTION: Ribozymes with broad tRNA aminoacylation activity
FILE REFERENCE: 11520 0290
CURRENT FILING DATE: 2003-02-18
PRIOR APPLICATION NUMBER: 60/357,424
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 61
SEQ ID NO 3
LENGTH: 74
TYPE: RNA
ORGANISM: artificial sequence
FEATURE:
NAME/KEY: misc_RNA
LOCATION:
OTHER INFORMATION: synthesized
US-10-369-036B-3

Query Match 25.3%; Score 38; DB 16; Length 74;
Best Local Similarity 82.4%; Pred. 0.00036;
Matches 61; Conservative 0; Mismatches 0; Indels 13; Gaps 1;

QY 90 GUGUGUAUCCCAAGGAGGUAACCGGACCGGA-----CAUUCGAGAUUCCGAUCC 136
DB 1 GUGUGUAUCCCAAGGAGGUAACCGGACCGGAUUCGGAUUCGGAUUCGGAUUC 60

QY 137 UCGUACCCGACGCA 150
DB 61 UCGUACCCGACGCA 74

RESULT 2

```
US-10-369-036B-12
; Sequence 12, Application US/10369036B
; Publication No. US20030228593A1
; GENERAL INFORMATION:
; APPLICANT: Suga, Hiroaki et al.
; TITLE OF INVENTION: Ribozymes with broad tRNA aminoacylation activity
; FILE REFERENCE: 11520.0290
; CURRENT APPLICATION NUMBER: US/10/369,036B
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,424
; NUMBER OF SEQ ID NOS: 61
; SEQ ID NO 12
; LENGTH: 56
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: synthesized
US-10-369-036B-12

Query Match      23.6%; Score 35.4; DB 16; Length 56;
Best Local Similarity 70.3%; Pred. No. 0.0032;
Matches 26; Conservative 10; Mismatches 1; Indels 0; Gaps 0;

OY      53 AUGGUAUUGCGGUAGUGCGGAGUAGUCUACGCU 89
      |||||:|||||:|||||:|||||:|||||:
Db      16 AAGGTAATTGGCGTTAGTGGCGGAGTGTCTACCT 52

RESULT 3
US-10-369-036B-8
; Sequence 8, Application US/10369036B
; Publication No. US20030228593A1
; GENERAL INFORMATION:
; APPLICANT: Suga, Hiroaki et al.
; TITLE OF INVENTION: Ribozymes with broad tRNA aminoacylation activity
; FILE REFERENCE: 11520.0290
; CURRENT APPLICATION NUMBER: US/10/369,036B
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,424
; NUMBER OF SEQ ID NOS: 61
; SEQ ID NO 8
; LENGTH: 101
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 45-48, 55-69
; OTHER INFORMATION: synthesized; n is g, a, t or c
US-10-369-036B-8

Query Match      22.0%; Score 33; DB 16; Length 101;
Best Local Similarity 75.8%; Pred. No. 0.03;
Matches 25; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

OY      1 GAUGGUAUGGUAUUGAGAUUUCGCGAGGCC 33
      |||||:|||||:|||||:|||||:|||||
Db      1 GGAATGCAATGTCATTGATTTCCGCGAGGCC 33

RESULT 4
US-10-369-036B-46
; Sequence 46, Application US/10369036B
; Publication No. US20030228593A1
; GENERAL INFORMATION:
; APPLICANT: Suga, Hiroaki et al.
; TITLE OF INVENTION: Ribozymes with broad tRNA aminoacylation activity
; FILE REFERENCE: 11520.0290
; CURRENT APPLICATION NUMBER: US/10/369,036B
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,424
; PRIOR FILING DATE: 2002-02-15
```

```
; NUMBER OF SEQ ID NOS: 61
; SEQ ID NO 46
; LENGTH: 77
; TYPE: RNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: synthesized
US-10-369-036B-46

Query Match      19.9%; Score 29.8; DB 16; Length 77;
Best Local Similarity 93.9%; Pred. No. 0.45;
Matches 31; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      53 AUGGUAUUGCGGUAGUGCGGAGUAGUCUACUA 85
      |||||:|||||:|||||:|||||:|||||
Db      27 AAGGUAUUGCGGUAGUGCGGAGUAGUCUACA 59

RESULT 5
US-10-369-036B-1
; Sequence 1, Application US/10369036B
; Publication No. US20030228593A1
; GENERAL INFORMATION:
; APPLICANT: Suga, Hiroaki et al.
; TITLE OF INVENTION: Ribozymes with broad tRNA aminoacylation activity
; FILE REFERENCE: 11520.0290
; CURRENT APPLICATION NUMBER: US/10/369,036B
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,424
; NUMBER OF SEQ ID NOS: 61
; SEQ ID NO 1
; LENGTH: 57
; TYPE: RNA
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION:
; OTHER INFORMATION: synthesized
US-10-369-036B-1

Query Match      19.6%; Score 29.4; DB 16; Length 57;
Best Local Similarity 96.8%; Pred. No. 0.6;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      53 AUGGUAUUGCGGUAGUGCGGAGUAGUCUAC 83
      |||||:|||||:|||||:|||||:|||||
Db      27 AAGGUAUUGCGGUAGUGCGGAGUAGUCUAC 57

RESULT 6
US-10-369-036B-2
; Sequence 2, Application US/10369036B
; Publication No. US20030228593A1
; GENERAL INFORMATION:
; APPLICANT: Suga, Hiroaki et al.
; TITLE OF INVENTION: Ribozymes with broad tRNA aminoacylation activity
; FILE REFERENCE: 11520.0290
; CURRENT APPLICATION NUMBER: US/10/369,036B
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,424
; NUMBER OF SEQ ID NOS: 61
; SEQ ID NO 2
; LENGTH: 76
; TYPE: RNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: synthesized
US-10-369-036B-2

Query Match      19.5%; Score 29.2; DB 16; Length 76;
Best Local Similarity 81.0%; Pred. No. 0.76;
```



```
US-10-027-632-119269/c
; Sequence 119269, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119269
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-119269

Query Match      19.3%; Score 29; DB 16; Length 1143;
Best Local Similarity 47.8%; Pred. No. 1.6;
Matches 33; Conservative 11; Mismatches 25; Indels 0; Gaps 0;

Qy      36 CUCACUACGCGUGGUCAGUAGUGGCGGAGUGGACUACGCGUGUGU 95
Db      1021 CACATATCTTGCGTTCACTGCTATGCGCTTGAGCGGAGTGCCTTCACTGCGGCG 962
Qy      96 AUCCCCAAG 104
Db      961 TGCCCAATG 953

RESULT 11
US-10-027-632-119270/c
; Sequence 119270, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119270
```

```
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-119270

Query Match      19.3%; Score 29; DB 16; Length 1143;
Best Local Similarity 47.8%; Pred. No. 1.6;
Matches 33; Conservative 11; Mismatches 25; Indels 0; Gaps 0;

Qy      36 CUCACUACGCGUGGUCAGUAGUGGCGGAGUGGACUACGCGUGUGU 95
Db      1021 CACATATCTTGCGTTCACTGCTATGCGCTTGAGCGGAGTGCCTTCACTGCGGCG 962
Qy      96 AUCCCCAAG 104
Db      961 TGCCCAATG 953

RESULT 12
US-10-027-632-119271/c
; Sequence 119271, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119271
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-119271

Query Match      19.3%; Score 29; DB 16; Length 1143;
Best Local Similarity 47.8%; Pred. No. 1.6;
Matches 33; Conservative 11; Mismatches 25; Indels 0; Gaps 0;

Qy      36 CUCACUACGCGUGGUCAGUAGUGGCGGAGUGGACUACGCGUGUGU 95
Db      1021 CACATATCTTGCGTTCACTGCTATGCGCTTGAGCGGAGTGCCTTCACTGCGGCG 962
Qy      96 AUCCCCAAG 104
Db      961 TGCCCAATG 953

RESULT 13
US-10-037-270-761/c
; Sequence 761, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
```



```

APPLICANT: Ren, Feiyun
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Tillinghast, John
APPLICANT: Dermanac, Radoje T.
TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
FILE REFERENCE: Polypeptides
CURRENT APPLICATION NUMBER: US/10/037,270
CURRENT FILING DATE: 2002-01-04
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: pc_Fl_genes Version 1.0
SEQ ID NO 761_
LENGTH: 2050
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (234)..(1445)
US-10-037-270-761

```

```

RESULT 14
US-10-117-722-761/c
? Sequence 761, Application US/10117722
? Publication No. US20030219744A1
? GENERAL INFORMATION:
? APPLICANT: Tang, Y. Tong
? APPLICANT: Liu, Chenghua
? APPLICANT: Asundi, Vinod
? APPLICANT: Zhang, Jie
? APPLICANT: Drmanac, Radoje T.
? TITLE OF INVENTION: No. US20030219744A1 Nucleic Acids and
? FILE OF INVENTION: Polypeptides
? FILE REFERENCE: 784CIP2BCIP
? CURRENT APPLICATION NUMBER: US/10/117,722
? CURRENT FILING DATE: 2002-04-04
? PRIOR APPLICATION NUMBER: 09/620,312
? PRIOR FILING DATE: 2000-07-19
? PRIOR APPLICATION NUMBER: 09/552,317
? PRIOR FILING DATE: 2000-04-25
? PRIOR APPLICATION NUMBER: 09/488,725
? PRIOR FILING DATE: 2000-01-21
? NUMBER OF SEQ ID NOS: 1104
? SOFTWARE: pc_FL_genes Version 1.0
? SEQ ID NO 761
? LENGTH: 2050
? TYPE: DNA
? ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (234) .. (1445)
US-10-117-722-761

```

	Query Match	19.2%	Score 28.8	DB 16	Length 2050
	Best Local Similarity	43.3%	Pred. No. 2.2		
	Matches	52	Conservative 11	Mismatches 57	Indels 0
				Gaps	0
Qy	30	GCCTTCACUAAACGUGGUCAUAGGUAUUGGCGUAGCGGAGUCUACGCU	89		
Db	206	GCCCCCTCCCCGGAGCCGGGGAATCCGCTGTGCTTAACTGCTCGATGCTCCACATGC	147		
Qy	90	GAUGUAUCCCAAGAGGUAACGGACCCGACAUUCGUAUUGAAUCCUCGUACCGCAGCC	149		
Db	146	TTGCGTCCACAGAAATGTCGCGCTTACGCCCGCGTTAGATCGAGTCCCGTAGCCGCTGCC	87		

```

RESULT 15
US-10-369-036B-49
: Sequence 49, Application US/10369036B
: Publication No. US20030228553A1
: GENERAL INFORMATION:
: APPLICANT: Suga, Hiroaki et al.
: TITLE OF INVENTION: Ribozymes with broad tRNA aminoacylation activity
: FILE REFERENCE: 11520.0290
: CURRENT APPLICATION NUMBER: US/10/369,036B
: CURRENT FILING DATE: 2003-02-18
: PRIOR APPLICATION NUMBER: 60/357,424
: PRIOR FILING DATE: 2002-02-15
: NUMBER OF SEQ ID NOS: 61
: SEQ ID NO 49
: LENGTH: 35
: TYPE: RNA
: ORGANISM: artificial sequence
: FEATURE:
: OTHER INFORMATION: synthesized
US-10-369-036B-49

```

Query Match	19.1%	Score	28.6	DB	16	length	35
Best local similarity	88.6%	Pred.	No. 1,1				
Matches	31	Conservative	0	Mismatches	4	Indels	0
QY	116	GGACATUUCGAGAUTUCCUGUACCGCGACCA	150				
DB	1	GAUGGACGAGAUTUCCUGUACCGCGACCA	35				

Search completed: October 8, 2004, 04:56:49
Job time : 402.667 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 7, 2004, 22:39:44 ; Search time 2909.33 Seconds
(without alignments)
1539.640 Million cell updates/sec

Title: US-09-721-414c-9

Perfect score: 150
Sequence: 1 ggaucgucagcagcagcagca.....ggaucgucagcagcagcagca 150

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hrc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hrc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	22.0	360	14	N41205 HO934-F Adu
2	32.8	21.9	875	12	B1226043 602950506
3	31.6	21.1	642	28	B2421617 hz32c07.b
4	31.2	20.8	857	29	CNS06X4M AL419276 T3 end of

Result No.	Score	Query Match	Length	ID	Description
5	31	20.7	777	12	B1818667
6	30.6	20.4	800	10	BE905814
7	30.6	20.4	932	29	CNS03P44
8	30.2	20.1	282	14	CK273925
9	30	20.0	405	10	BE429745
10	29.8	19.9	576	28	A2876215
11	29.8	19.9	688	14	CF442540
12	29.8	19.9	1101	29	CNS003EW
13	29.8	19.9	1246	13	B0433823
14	29.4	19.6	409	10	BF187374
15	29.4	19.6	418	12	BI920370
16	29.4	19.6	470	12	CB106726
17	29.4	19.6	506	14	BI178400
18	29.4	19.6	535	13	B0119414
19	29.4	19.6	536	12	B1406735
20	29.4	19.6	552	12	B1432404
21	29.4	19.6	559	12	BS689741
22	29.4	19.6	555	13	B0120597
23	29.4	19.6	564	12	BM111158
24	29.4	19.6	565	12	BS692017
25	29.4	19.6	568	12	B1433590
26	29.4	19.6	570	12	BQ047394
27	29.4	19.6	571	14	CB104808
28	29.4	19.6	572	14	CK268944
29	29.4	19.6	575	12	BG098763
30	29.4	19.6	589	12	BK404648
31	29.4	19.6	592	12	BM112684
32	29.4	19.6	593	10	BE924492
33	29.4	19.6	595	12	BQ046009
34	29.4	19.6	603	12	BM109181
35	29.4	19.6	611	12	BS689522
36	29.4	19.6	612	12	BM113610
37	29.4	19.6	628	10	BF051109
38	29.4	19.6	638	12	BQ045762
39	29.4	19.6	651	12	BM113345
40	29.4	19.6	664	12	BK406770
41	29.4	19.6	671	12	BS691282
42	29.4	19.6	671	12	BM110720
43	29.4	19.6	676	13	BQ114812
44	29.4	19.6	676	12	BG594714
45	29.4	19.6	677	12	BM405539

ALIGNMENTS

RESULT 1
N41205/c
LOCUS
DEFINITION HO934-F Adult heart, Clontech Homo sapiens CDNA clone Q934-F, mRNA
ACCESSION N41205
VERSION N41205.1 GI:1164803
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
Wu, M.Y., Cheung, H.K.Y., Lam, W.Y., Law, P.T.W., Lo, A.S.Y., Liu, V.W.Y., Luk, S.C.W., Tsui, S.K.W., Tung, C.K.C., Yam, N.Y.H., Liew, C.C. and Lee, C.Y.
Gene expression of adult human heart as revealed by random sequencing of cDNA library
Miami Winter Biotechnol. Symp. Proc. 6, 90 (1995)
Contact: Wayne Mary M.Y.
Department of Biochemistry
The Chinese University of Hong Kong
Rm 302C, Basic Medical Science Building, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.
Tel: 8526096874
Fax: 8526035123
Email: bl33723@vax.csc.cuhk.hk

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

FEATURES
SOURCE

ORIGIN

Query Match
Best Local Similarity 43.6%; Pred. No. 1,3e+02;
Matches 44; Conservative 13; Mismatches 44; Indels 0; Gaps 0;

Oy 45 GGUGGUCUAGCGGUAVUNGGCGGUAAGUGCAGACGCGUGGCUAUGCCCAAG 104
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 422 GTGGGTGTTTGCTGGTGTCTTATGTAAGATTGACAACCATGCAGCTGGCTGTAAAC 363
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Oy 105 GGUACGGGACCAGCAUUDCGAUUDCGAUUDCCUGUACCGC 145
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 362 GAGCCGTGCGCCGAGAAGCGGCTTCACACTGCATGCCTC 322
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 8
CK273925/c LOCUS 282 bp mRNA linear EST 12-DEC-2003
DEFINITION E57720003 potato abiotic stress cDNA library Solanum tuberosum CDNA
clone POADE22 3' end, mRNA sequence.
ACCESSION CK273925
VERSION CK273925.1 GI:39830903
KEYWORDS EST.

SOURCE ORGANISM Solanum tuberosum (potato)
Solanum tuberosum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
1 (bases 1 to 282)
Buell,C.R., Hart,A., Zisemann,V., Karanycheva,S.A. and Baker,B.
Generation of ESTs from abiotic stressed potato tissue
Unpublished (2003)
Other ESTs: E57720002
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potaco-array@tigr.org
Clones can be requested from TIGR via potaco@tigr.org
Seq primer: GTA ATA CGA CTC ACT ATA GGG C.

REFERENCE AUTHORS
TITLE JOURNAL
COMMENT

FEATURES
SOURCE

Location/Qualifiers
1..282
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="POADE22"
/tissue type="abiotic stress treated leaf and root tissue"
/lab host="DH10B-Tona"
/clone.lib="potato abiotic stress cDNA library"
/note="Vector: pCMVSPORT6.1; Site_1: EcoRI; Site_2: NotI;
supplier: Solanum tuberosum var. Kennebec plants were

with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

	Query Match	Similarity	Score	DB	Length
Best Local	50%	19.9%	29.8	28	576
Matches	Conservative	44.2%	Pred. No. 2e+02		
		11	Mismatches	52	Indels
					Gaps
Qy	7	UcAGUGcAnUgAGuUuCCGcAGcGccUuCUcAUCuAGuGcGcAUGGAGUuGcGU	66		
Db	73	TcGcGcCtTgGgGAGtCtGcTcGcGcCtTcGcAAGAcAcTcGcGcGcGcTtCAcTcGcG	132		
Qy	67	UAGUGcGcGAGUcUcAUCUcAGcUGuGAGUuCCCAAGcGUAcGgAGcCGAc	119		
Db	133	TAcGtGgGcGtTtTgGcCCGcGcAcAGcGcAcAGtCtTcAGGcGcGcGcGcTcTcGcGcG	185		

RESULT 11

LOCUS	CP442540	688 bp	mRNA	linear	EST 04-SEP-2003
DEFINITION	EST678885 normalized cdna library of onion Allium cepa cDNA clone				
ACCESSION	CP442540				
VERSION	CP442540.1	GI:34465242			
KEYWORDS	EST.				
SOURCE	Allium cepa (onion)				
ORGANISM	Allium cepa				

REFERENCE	1 (bases 1 to 688)
AUTHORS	Havey,M.J., Cheung,F., Van Aken,S., Uitterback,T. and Town,C.D.
TITLE	Expressed Sequence Tags from a normalized library of mixed oncon tissues (Allium cepa)
JOURNAL	Unpublished (2003)
COMMENT	Contact: Havey MJ

FEATURES

```

1. 688
/organisms="Allium cepa"
/mol_type="mRNA"
cultivar="Red Creole (bulbs), unknown (callus), Ebano &
Texas Legend (roots)"
/db_xref="taxon:4679"
/clone="ACAFV13"
/tissue_type="Callus, roots, and young bulbs"
/clone_lib="normalized cDNA library of onion"
/notes="Vector: pCMSPORT6.1-cdb (Invitrogen); Site 1:
EcoRI (5'); Site 2: NotI (3'); Equal molar amounts of
from callus, roots, and young bulbs were combined to
synthesize the library. Normalization to enrich for
low-copy transcripts was performed by proprietary
techniques of Invitrogen."

```

ORIGIN

Query Match	19.9%	Score 29.8	DB 14	Length 688
Best Local Similarity	50.7%	Pred. No. 2	le=08	
Matches	37	Conservative	9	Mismatches 27
				Indels 0
				Gaps 0

Qy	62	GCGGUAAGUGCGGAGUCACUACGCTGUGGUGUAUCCCCAGGAGGUAAGGACGGACAT	121
Db	658	GAGCGTTCCTCGGGGCCCGCTAAATGAGCTGGTGGAAAGGCAAGCCACATCGGCCCGGAAAT	599

QY	122	UCGAGAUUCCGAU	134
		: :	
Db	598	TGCGTCTCCGAGT	586

RESULT 12

LOCUS CENS033EM 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
BAOR08K02 of RPECI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL064622
VERSION AL064622.1 GI:4941974
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)

REFERENCE

TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a

FEATURES

```
1.1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clona="BACR08K02"
/clone_1fb="RPCI-98"
/notes="end : T7"
```

ORIGIN

		19.9%;	Score 29.8;	DB 29;	Length 1101;
		Best Local Similarity	37.1%;	Pred. No. 2.4+02;	
Matches	39;	Conservative	22;	Mismatches	44; Indels 0; Gaps 0;
Cy	45	GGUGGUAUCAGUAGUUUGCGUGUGGCGGAUACUACGUGUGGUAUCCCCAAG	104		
		:::::::::::::::	:	:	:
Dd	1022	GGRKACCCTGAAATTKGSKTTTAAAAGYAAAAAKRTTTCMCCCGARARAAAACCSGA	963		
		:::::::::::::::	:	:	:
Oy	105	GGUACGGGACCGGACAUUCCGAGUUDCGAUUTCUGUACCGCAGCC	149		
		:::::::::::::::	:	:	:
Dd	962	AAAAACAATAAAAGGTGTCGAAATAAACCTCMGTGCGCCSCC	918		

RESULT 13

LOCUS	BQ433823	1246 bp	mRNA	linear	EST 24-MAY-2007
DEFINITION	AGNCCOCT 7580839 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5046784 5', mRNA sequence.				
ACCESSION	BQ433823				
VERSION	BQ433823.1	GI:21172899			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				


```

/dev stage="7", 8 and 10 days"
/lab_host="RSOLR"
/clone_lib="potato microtubers, in vitro-grown"
/notes=Vector: plusscript SK(-); Site_1: ECORI; Site_2:
XhoI; Tissue supplied by Christian Bache and Richard
Visser (Department of Plant Breeding, Wageningen
University, The Netherlands); sequencing by The Institute
for Genomic Research. The cDNA libraries will attempt to
capture the induction and initiation/initial growth of the
tuber in an in vitro system as described in Bache et al.
(Plant Journal, 1996). Small microtubers develop from
axillary buds attached to stem explants when placed on a
high sucrose medium (10%). Visible morphological changes
occur synchronously at day five in the axillary buds. The
first library, cDNA (1-20) consists of axillary buds
harvested on days 1-3. This targets those genes involved
in induction of the microtubers. The following libraries,
cDNA (21-40) and cDNA (41-60), capture genes involved in
tuber initiation and outgrowth. This library is noted as
p3 in Tanksey lab notebooks."

```

ORIGIN

Query Match	19.6%	Score 29.4	DB 12	Length 418
Best Local Similarity	46.3%	Pred. No. 2.4e+02		
Matches 44	Conservative 10	Mismatches 41	Indels 0	Gaps 0

Oy AACGGUGGUCANUGGUAUUGGCAGUAAGGUGCGGAUGVCAACUACGUGGUAUCC 101
42 | C G G U G G U C A N U G G U A U U G G C A G U A A G G U G C G G A U G V C A A C U A C G U G G U A U C C |
Db AACGGTACTTCATCATGGAATGGGGCTATGGTACGGCGCACACAATACTCATGGGATGGC 359
300 | A A C G G T A C T T C A T C A T G G A A T G G G G C T A T G G T A C G G C G C A C A C A A T A C T C A T G G G A T G G C |

Oy AAAGGUACGGGACCGGACAUUUGAGAUAUCC 136
102 | A A G G U A C G G G A C C G G A C A U U U G A G A U A U C C |
Db TAATGTACGGGCCAACAACTAACAACCTTAAGCC 394
360 | T A A T G T A C G G G C C A A C A A C T A A C A A C C T T A A G C C |

Search completed: October 8, 2004, 04:06:03
Job time : 2913.33 secs